



Punishment sensitivity and risk taking in depressed mood

Dahyeon Kim¹ · Jovian Lam¹ · Amanda Kutz² · K. Lira Yoon^{1,3}

Accepted: 2 December 2020 / Published online: 2 January 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

Individuals with high levels of depressive symptoms find rewards to be less reinforcing and punishment more aversive, which may lead them to avoid risk taking. To examine the relation between risk taking and depressive symptoms, undergraduate participants ($N = 138$) completed the Balloon Analogue Risk Task (BART), which effectively measures affective risk taking. Participants' goal was to maximize the virtual money earned from the task by completing trials that ended in either a reward (i.e., gain in virtual money) or a punishment (i.e., loss of virtual money and premature end to the trial). Higher (vs. lower) levels of depressive symptoms were associated with more pumps (i.e., greater risk taking) after a reward; higher levels of depressive symptoms were associated with fewer pumps following a punishment than a reward (i.e., greater punishment sensitivity). Greater punishment sensitivity (i.e., lower risk-taking following setbacks) in individuals with higher levels of depressive symptoms may contribute to the maintenance of depressed mood by limiting rewarding experiences that could elevate their mood. The findings should be replicated in a clinical sample.

Keywords Depression · Risk taking · Punishment sensitivity · Decision making · Reward sensitivity

Introduction

Although risk taking typically has a negative connotation (e.g., pathological gambling, drug abuse), it can also be advantageous (i.e., positive risk taking). Specifically, risk taking can be beneficial when the behavior is socially acceptable, punishment is relatively unlikely, and rewards are potentially significant (e.g., applying for a job). Nevertheless, individuals with higher levels of depressive symptoms avoid taking risks, including positive ones (e.g., participating in clinical trials; Leykin et al. 2018). In turn, this avoidance of risk taking may limit the exposure to situations that challenge pre-existing negative schemas, thereby

maintaining depression (Beck 2008). Different sensitivity to rewards and punishments may drive avoidance of risk taking in depression. Considering that a negative schema in depressed mood facilitates the processing of negative information (e.g., punishment) and impedes the processing of positive information (e.g., rewards) (Beck 2008), avoidance of risk taking in depression could arise from low reward sensitivity, high punishment sensitivity, or the combination of both.

Rewards shape behavior by increasing the behavior it follows; smaller physiological, emotional, or behavioral changes following a reward indicate reduced reward sensitivity. Thus, lower risk taking following rewards in depression reflects reduced reward sensitivity, likely related to avolition and anhedonia (Henriques and Davidson 2000). Indeed, both individuals diagnosed with depression (Pizzagalli et al. 2008) and individuals with higher levels of subclinical depressive symptoms (Pizzagalli et al. 2005) select the more rewarded option less frequently than healthy controls. Further, unlike healthy controls who took more risks during trials with rewards than trials without rewards, individuals with higher levels of depressive symptoms did not alter their behavior depending on the availability of rewards (Henriques and Davidson 2000).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11031-020-09860-4>.

✉ Dahyeon Kim
dkim22@nd.edu

¹ Department of Psychology, University of Notre Dame, 390 Corbett Family Hall, Notre Dame, IN 46556, USA

² VA Salt Lake City Health Care System, Salt Lake City, UT, USA

³ Department of Psychology, University of Maryland Baltimore County, Baltimore, MD, USA

Punishment sensitivity may also lead individuals with higher levels of depressive symptoms to avoid risk taking. Based on cognitive theories of depression (e.g., Beck 2008), heightened sensitivity to punishments, as indicated by greater behavioral, emotional, cognitive, or physiological change in anticipating or following a punishment, is expected in depression. Indeed, when required to solve a puzzle in the fewest steps possible, individuals with depression made more errors following an error (Holmes and Pizzagalli 2008; Steffens et al. 2001). To investigate risk taking specifically, another study (Smoski et al. 2008) used the Iowa Gambling Task (IGT; Bechara et al. 1994), which requires individuals to learn contingencies of rewards and punishments. In this study, Smoski et al. found that individuals with depression (vs. healthy controls) took fewer risks. These findings suggest that higher levels of depressive symptoms are associated with greater sensitivity towards punishment. Other studies, however, found no difference in punishment sensitivity between individuals with depression and healthy controls. For example, two studies (Deisenhammer et al. 2018; Gorlyn et al. 2013) also used the IGT and found that individuals with depression took similar levels of risk as healthy controls. Therefore, despite theoretical and some empirical evidence suggesting heightened punishment sensitivity in depression, the association between punishment sensitivity and depression remains unclear.

Tasks used in previous studies may have contributed to these inconsistencies. One of the most frequently used task is the IGT, in which individuals are asked to choose cards from different decks that are either “bad” (i.e., high-reward, high-punishment) or “good” (i.e., low-reward, low-punishment). Healthy individuals tend to learn over time to choose more from the “good” decks while avoiding the “bad” decks. Importantly, risk taking in the IGT involves not only emotional decision-making, but also rational decision-making (e.g., risk perception, deliberate evaluation of alternatives) (Buelow and Blaine 2015). Thus, risk taking in the IGT is also influenced by impaired executive functioning and other cognitive processes (e.g., Rock et al. 2014).

The Balloon Analogue Risk Task (BART; Lejuez et al. 2002) may be a purer measure of the affective aspect of risk taking than the IGT, and thus, better assess punishment sensitivity in depressed mood (Buelow and Blaine 2015). In the traditional BART, participants pump a virtual balloon to maximize total earned virtual money. On each trial, pumping the balloon (i.e., risk taking) may either increase the virtual money earned (i.e., rewarded) or burst the balloon, the latter of which would end the trial and result in the loss of the money earned in that trial (i.e., punished). In the IGT, participants learn the strategic approach throughout trials. Specifically, they accumulate knowledge about the payoffs associated with different decks across different trials. Therefore, performance—and by extension, punishment

sensitivity—on the IGT is dependent on the participants’ learning speed, which is not the case in the BART. Furthermore, in the BART, each pump that individuals decide (not) to administer is an opportunity for taking risks, presented multiple times per trial. This high frequency of risk-taking opportunities makes the traditional BART an ideal tool for assessing punishment sensitivity. Despite these advantages, research examining the relation between punishment sensitivity in BART and depressive symptoms is limited.

Importantly, the traditional BART has been criticized for its bias for lower number of pumps (i.e., lower risk taking) (Hevey et al. 2017; Pleskac et al. 2008). Performance on the traditional BART is often analyzed using the average number of pumps across all trials (i.e., unadjusted pumps) or the average number of pumps on trials that did not end with a burst (i.e., adjusted pumps), using an analysis of variance (ANOVA). These scores tend to be low, because it is less likely for a trial with fewer pumps to end in a burst compared to a trial with more pumps (Pleskac et al. 2008).

A modified version, the automatic BART (Pleskac et al. 2008), attempts to address this low score bias by having participants enter their intended number of pumps at the start of each trial. This way, the number of pumps administered is unaffected by the burst of a balloon. Using the automatic BART, one study (Hevey et al. 2017) found heightened punishment sensitivity in depression (i.e., lower risk taking following a loss). Specifically, the change in the average number of pumps from successful trials (i.e., ending in a gain) to unsuccessful trials (i.e., ending in a balloon burst) was significantly greater in individuals with depression, compared to healthy controls. Although the automatic BART addresses the low score bias, critical information may be lost. Because the automatic BART eliminates the need to pump the balloon manually, risk taking opportunities are greatly reduced (i.e., the participant takes one risk per trial when they enter the number of pumps, rather than taking a new risk with each successive pump). Relatedly, the emotion-inducing situation that occurs each time a participant takes risks (i.e., decides whether to pump the balloon) is absent in the automatic BART. In addition, the automatic BART informs explicitly the optimal number of pumps per trial, introducing rational decision-making components. Therefore, performance on the automatic BART may be less affected by emotions associated with rewards and punishments and more by logic than on the traditional BART.

Using multilevel modeling (MLM) allows us to take advantage of the critical elements of the traditional BART while addressing its low score bias. MLM has several benefits over ANOVA, especially when examining BART performance. By nesting trials within each individual and controlling for the presence of burst in each trial, MLM differentiates the number of pumps in prematurely ended trials from trials that participants chose to

end. Additionally, considering the trial-to-trial variance by examining the trial-level data, instead of averaging across trials, ensures greater power to detect the effects of interest. It also addresses the inflation of Type I error that results from categorizing the continuous variable of trials into one average performance measure per individual necessary in ANOVA (Maxwell and Delaney 1993). Thus, MLM has been identified as a powerful tool in analyzing data from experimental designs with repeated measurements (Hoffman and Rovine 2007).

In the present study, we aimed to clarify the relation between depressive symptoms and risk taking following punishments (i.e., punishment sensitivity). To this end, we analyzed performance on the traditional BART using MLM, the task and the statistical procedure selected to address possible confounds from logical decision-making components of previously used tasks. Based on previous findings that demonstrated less risk taking (Smoski et al. 2008) and heightened punishment sensitivity (Hevey et al. 2017) in individuals with high levels of depressive symptoms, we hypothesized that individuals with higher (vs. lower) levels of depressive symptoms would pump the balloon fewer times (i.e., avoid risk taking), especially after a punishment trial (i.e., display heightened punishment sensitivity).

Methods

Participants

A total of 153 undergraduates participated in the study for course credit. Mean age was 19.78 ($SD=4.22$), and 76% of the participants were female.

Measures

Center for epidemiology survey-depression

The Center for Epidemiology Survey-Depression (CES-D; Radloff 1977) is a 20-item measure that assesses the level of depressive symptoms for the general population. The CES-D better discriminates the severity of depressive symptoms in college students than the Beck Depression Inventory (Santor et al. 1995). Although the suggested cut-off score might overestimate depression in a nonclinical sample, the CES-D has exhibited superior discriminability in detecting differences in depressive symptom severity among college students (Santor et al. 1995). It also has good internal

consistency between 0.85 and 0.90 and adequate test–retest reliability (Pearson's r) between 0.51 and 0.67 (Radloff 1977). In the current study, depressive symptoms in the past two months were assessed, and Cronbach's alpha was 0.72.

Balloon analogue risk task

The Balloon Analogue Risk Task (BART; Lejuez et al. 2002) is a reliable and valid measure of risk taking. For example, BART performance translated to real-world risk-taking behavior, such as smoking, in a college sample (Lejuez et al. 2003). We used a version programmed in the Psychology Experiment Building Language (PEBL; Mueller and Piper 2014). Participants pressed a button to pump a virtual balloon presented on the computer screen. Each pump earned the participants five virtual cents. Each trial ended when the participant chose to do so or pumped the balloon more than the number of pumps programmed, bursting the balloon. If the participant chose to end the trial, the money earned through the last pump was transferred to their permanent bank. If the balloon burst, the participant lost the money earned for that trial (see Fig. 1). Thus, pumping more (i.e., taking more risks) can result in gaining more (i.e., receiving reward) or losing the earned virtual money (i.e., being punished). The participants did not receive the value of the virtual money but were falsely informed that they would be entered in a raffle for a gift card if they performed higher than average. In truth, all participants were entered in this raffle.

The participants completed a total of 90 trials, divided into four phases. The participants were expected to learn the color-dependent burst thresholds during the Learning Phase (trials 1–30), in which ten orange, ten yellow, and ten blue balloons were randomly presented. Next, the participants were presented with all orange, all yellow, and all blue balloons in Phase 1 (trials 31–50), Phase 2 (51–70), and Phase 3 (71–90), respectively. The threshold for pumps before bursting was the lowest for the orange balloons (average of four pumps and a range between one and eight pumps), highest for the blue (average 64, range = 1–128), and intermediate for the yellow (average 8, range = 1–16).

Procedure

Participants provided written informed consent. They completed the BART and other tasks unrelated to this study. Task order was counterbalanced for all participants to control for potential order effects. Following the tasks, participants completed questionnaires, including the CES-D, and were debriefed about the deception regarding the raffle. All procedures were approved by the local Institutional Review Board.

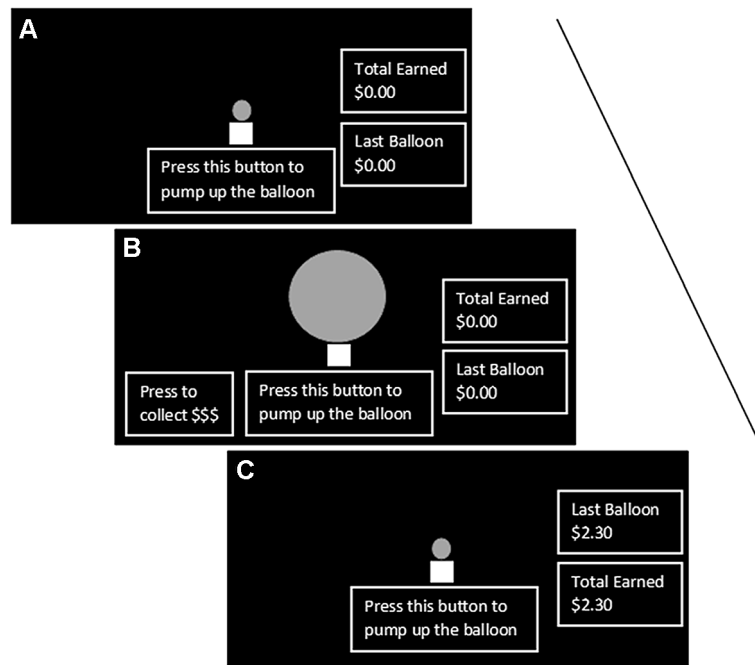


Fig. 1 The balloon analogue risk task (BART). *A* When a trial begins, the participant is presented with a visualization of a balloon, the total earned amount from the task, and the amount earned from the last trial; *B* if burst threshold is not met, each pump increases the size of the virtual balloon, and the participant can choose to end the trial by pressing the collect button. If the burst threshold is met, the balloon disappears to represent the burst; *C* if the participant presses the col-

lect button, the amount earned on the trial is presented next to “Last Balloon.” The participants are also presented the total earned, which is the sum of the “Last Balloon” and the amount that was presented as the total earned in the previous trial. If the balloon bursts, the amount earned on that trial is presented as \$0.00, and the total earned remains unchanged from the previous trial

Table 1 Means (SDs) of participant characteristics and task performance

	N	Mean	SD
CES-D	138	21.99	6.91
Phase 1 pumps	138	2.55	0.71
Phase 1 virtual money collected (\$)	138	1.05	0.34

CES-D center for epidemiology survey-depression

Results

Participant characteristics

The participant characteristics and performance on the BART are presented in Table 1. Of 153 participants, 15 participants had missing data (eight missing the BART, four missing the CES-D, and three missing both), thereby resulting in a total of 138 participants. Mean age was 19.81 ($SD = 4.33$) for the final sample; 77.4% of the participants identified as female, 22.6% as male, and one participant chose not to respond. With multiple group identification

permitted, 87.7% of the participants identified as Caucasian, 2.9% as Asian, 2.9% as Native, 2.2% as Black American; and 5.1% identified as “other.” Considering the recommended cut-off score of 16 to indicate mild levels of depression (Shean and Baldwin 2008), our sample exhibited higher levels of depressive symptoms than expected in a college sample (see Table 1).

Learning phase

Participants’ performance during the Learning Phase was analyzed to ensure that they learned the contingencies associated with the balloon colors. Therefore, a multilevel model was fitted, including the balloon color (orange, yellow, balloon) \times time (continuous) as a level-1 predictor. Controlling for the previous punishment, current trial punishment (level-1 predictors), and depressive symptoms (level-2 predictor), the main effect of time was significant, $t(137) = 3.89, p < 0.001$. Individuals took fewer risks (i.e., pumped fewer times) as they progressed through the phase ($\beta = -0.01, SE = 0.002$). The main effect of the balloon color was also significant: The participants took more risks for colors associated with higher thresholds

($\beta = 0.25$, $SE = 0.02$), $t(137) = 12.88$, $p < 0.001$. Importantly, the time \times balloon color interaction was significant, $t(136) = 2.34$, $p = 0.02$, indicating that the participants learned the correct contingencies over time ($\beta = 0.002$, $SE = 0.001$). No effect involving depressive symptoms was significant (all $ps > 0.24$), indicating that learning did not depend on levels of depressive symptoms.

Depression and BART performance

Phase differences

For testing the hypothesis, we expected Phase 1 data to be the most appropriate for two reasons: (1) The lower burst threshold for the orange balloons resulted in more bursts of balloons (i.e., more frequent punishment) and, thus, may be better suited to examine punishment sensitivity; and (2) lower threshold balloons may minimize the influence of motivation and boredom resulting from the tedious nature of manually administering pumps in the traditional BART. Specifically, participants might get bored and pump fewer times simply to end trials early. As a result, their performance on later phases might be affected by boredom and a lack of motivation rather than risk taking and punishment sensitivity.

To examine the speculation that the later phases may be more likely to be affected by boredom regardless of depressive symptoms, we conducted a phase (Phase 1, 2, and 3) \times depressive symptoms (a continuous variable) repeated-measures analysis of covariance (ANCOVA) on the proximity to burst threshold scores. The proximity to burst threshold scores were calculated by dividing the number of pumps participants administered by the median threshold for each balloon, thereby controlling for the differences in threshold by phase. A phase \times depressive symptoms interaction was not significant, $F(2, 135) = 1.39$, $p = 0.25$, and the main effect of depressive symptoms was also not significant, $F(1, 136) = 2.69$, $p = 0.10$. Importantly, participants pumped the balloon less with each consecutive phase, $F(2, 135) = 11.59$, $p < 0.001$, $\eta_p^2 = 0.15$, even though the threshold increased in the later phases. Thus, the number of pumps in the later phases might have been affected by factors other than general risk taking or punishment sensitivity (e.g., boredom).

We also examined whether there were fewer bursts in the later phases, which would suggest a decline in the participants' motivation to maximize their points. Thus, we conducted a phase (Phase 1, 2, and 3) \times depressive symptoms repeated-measures ANCOVA on the number of bursts. The main effect of phase was significant, $F(2, 135) = 9.74$, $p < 0.001$, $\eta_p^2 = 0.13$, and, as expected, there were fewer bursts in the later phases. The main effect of depressive symptoms was significant, $F(1, 136) = 5.45$, $p = 0.02$, $\eta_p^2 = 0.04$, such that individuals with higher

levels of depressive symptoms had more bursts overall. A phase \times depressive symptoms interaction was not significant, $F(2, 135) = 0.73$, $p = 0.48$ (see Supplementary Material for more information). Therefore, we used only the Phase 1 data to test the hypothesis, considering that it occurred earlier in the task and provided more trials ending in punishments compared to other phases. However, results from other phases are also available as Supplementary Material.

Main analysis (phase 1)

In Phase 1, the participants' depressive symptoms did not correlate with the number of pumps, $r(137) = 0.13$, $p = 0.13$, nor the proportion of trials ending in punishments, $r(137) = 0.16$, $p = 0.06$. To investigate the interactive effect of depressive symptoms and previous loss on risk taking (i.e., punishment sensitivity), we performed MLM on Phase 1 data. To measure risk taking, the log transformed number of pumps served as a continuous time-varying outcome variable. Whether a punishment trial preceded the current trial, a time-varying and binary variable (Level 1), was nested within individuals (Level 2) as a possible predictor of the number of pumps. If punishment did not precede the current trial, this preceding trial was considered a reward trial, due to the binary design of the task. The centered trial number and whether the current trial was a punishment trial (a binary predictor) were also nested as level-1 time-varying covariates. The following describes the Level 1 model:

$$\text{Pumps}_{it} = [b_{0i} + b_{1i} \times (\text{trial} - 1)_{it} + b_{2i} \times \text{punishment}_{it} + b_{3i} \times \text{punishment}_{it-1}] + [e_{it}]$$

For each participant i , the intercept (b_{0i}) represents the number of pumps at the start of the task. The b_{1i} represents the slope of the change in the number of pumps with each trial (t). The b_{2i} represents the slope of the change in the number of pumps in a current punishment trial. The b_{3i} represents the slope of the change in the number of pumps when the previous trial ($t - 1$) was a punishment trial.

The grand-mean centered level of depressive symptoms was included as a continuous time-invariant variable (Level 2). Given our interest in depressive symptoms, we focused on the fixed effect of the level-2 predictor. The following describes the Level 2 model:

$$b_{0i} = \beta_{00} + \beta_{01} \times \text{CES} - D_i + z_{0i}$$

$$b_{1i} = \beta_{10} + z_{1i}$$

$$b_{2i} = \beta_{20} + z_{2i}$$

$$b_{3i} = \beta_{30} + \beta_{31} \times \text{CES} - D_i + z_{3i}$$

Thus, the multilevel model with time-varying and time-invariant predictors, including the previous punishment (binary) \times depressive symptoms (continuous) as a level-2 predictor, was fitted.

The main effects of previous punishment, $t(135) = 2.64$, $p = 0.01$, and the level of depressive symptoms were significant, $t(2210) = 2.19$, $p = 0.03$. As expected, individuals displayed punishment sensitivity by pumping the balloon fewer times (i.e., avoiding risk taking) after a trial ending with a loss ($\beta = -0.03$, $SE = 0.02$). Contrary to our expectations, however, individuals with higher (vs. lower) levels of depressive symptoms took more risks ($\beta = 0.003$, $SE = 0.002$), controlling for the other predictors.

Importantly, these main effects were qualified by a significant previous punishment \times depressive symptom interaction, $t(2210) = 1.99$, $p = 0.047$ (see Fig. 2). Probing the interaction through the online tool provided by Preacher et al. (2006), we found that individuals with higher levels of depressive symptoms (i.e., 1 SD above the average) displayed greater punishment sensitivity ($\beta = -0.003$, $SE = 0.002$), such that they pumped significantly fewer times following punishments (vs. rewards). Specifically,

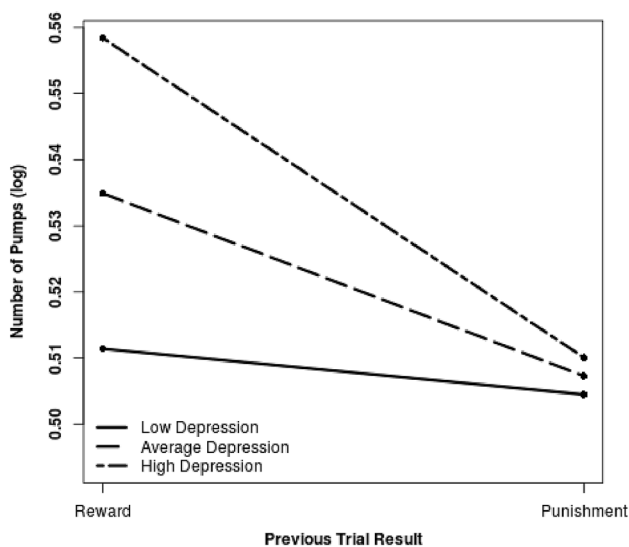


Fig. 2 Number of pumps (log transformed) as a function of depressive symptoms and the presence of punishment in a preceding trial. Following a previous punishment (vs. reward), individuals with higher levels of depressive symptoms (high depression) pumped the balloon significantly fewer times. Average depression = participants whose scores on the Center for Epidemiology Survey-Depression Scale (CES-D) were at the mean level. Low depression = 1 standard deviation below the mean CES-D score (i.e., individuals with lower levels of depressive symptoms). High depression = 1 standard deviation above the mean CES-D score

the number of pumps depended on whether the previous trials were punishments in participants with average, $t(2210) = 17.44$, $p < 0.001$, and higher, $t(2210) = 4.59$, $p < 0.001$, but not with lower, $t(2210) = 0.60$, $p = 0.55$, levels of depressive symptoms. We also found that individuals with higher (vs. lower) levels of depressive symptoms took significantly more risks in the trials following rewards, $t(2210) = 2.19$, $p = 0.03$, but not in the trials following punishments, $t(2210) = 0.27$, $p = 0.79$.

Discussion

We examined whether individuals with higher levels of depressive symptoms avoided risk taking, especially following a punishment, thereby displaying higher punishment sensitivity. Consistent with the hypothesis, higher levels of depressive symptoms were related to pumping fewer times (i.e., less risk taking) after a punishment trial than a reward trial, suggesting heightened punishment sensitivity. Unexpectedly, individuals with higher levels of depressive symptoms took more risks than individuals with lower levels of depressive symptoms after rewards; risk taking, however, did not depend on the severity of depressive symptoms following punishments. Meanwhile, risk taking in individuals with lower levels of depressive symptoms did not depend on whether they were punished or rewarded in the previous trial. In sum, individuals with higher levels of depressive symptoms demonstrated higher sensitivity for punishments compared to rewards; they also demonstrated higher sensitivity for rewards than individuals with lower levels of depressive symptoms.

Our result demonstrating greater behavioral changes in individuals with higher levels of depressive symptoms following a punishment (vs. rewards) mirrors findings of heightened punishment sensitivity in depression (Holmes and Pizzagalli 2008; Hevey et al. 2017; Smoski et al. 2008; Steffens et al. 2001). In previous studies, depressive symptoms were associated with behavioral changes that reflect heightened sensitivity to punishments, such as committing more errors following a failure to solve a puzzle (Holmes and Pizzagalli 2008; Steffens et al. 2001) and selecting more from options with smaller punishment (Smoski et al. 2008). In the current study, the heightened punishment sensitivity manifested as fewer pumps following a balloon burst than following virtual money collected. Broadly, the current finding suggests that more depressive symptoms are associated with fewer risk taking after a loss or punishment (vs. reward).

This heightened punishment sensitivity may be explained by the negative schemas, which drive avoidant behaviors (e.g., Beck 2008) and have been linked to avoiding risks (Leahy et al. 2012). Following a setback, individuals with higher levels of depressive symptoms may perceive punishments as more likely to occur than rewards, due to their greater pessimism (Roepke and Seligman 2016). In turn, this pessimism and their perception of higher likelihood of punishment may promote avoidant behaviors that minimize opportunities for rewards and positive affect (e.g., Carvalho and Hopko 2011), thereby reinforcing negative schemas. Thus, our results align with the cognitive-behavioral theories of depression (Beck 2008) that emphasize the cycle of dysfunctional negative schemas and behavioral avoidance in the maintenance of depression. The heightened punishment sensitivity may also be associated with the greater activation following a punishment, in individuals with depression than in healthy controls, of the medial prefrontal cortex, which is involved in self-monitoring and behavioral adaptations based on reinforcement contingencies (Cavanagh et al. 2011).

Unexpectedly, individuals with higher (vs. lower) levels of depressive symptoms exhibited greater risk taking following rewards. That is, individuals with higher (vs. lower) levels of depressive symptoms in our sample demonstrated heightened reward sensitivity, contradicting past research on reward sensitivity (e.g., Henriques and Davidson 2000). Although relatively more depressed, participants with more depressive symptoms in our sample of college students are likely highly functioning, as indicated by their ability to continue their studies and attend the lab session. Higher perfectionism, which has been associated with greater depressive symptoms (e.g., Hewitt and Flett 1991), might allow individuals with higher levels of depression to maintain their performance (i.e., the so-called high-functioning depression). Indeed, a previous study found higher levels of perfectionism in the college students with high levels of depressive symptoms, compared to both healthy controls and patients diagnosed with major depressive disorder (Enns et al. 2001). Higher levels of perfectionism in dysphoric individuals could lead to greater attempts and efforts to maximize their performance on a given task, resulting in the positive association between depressive symptoms and risk taking when reward is present (i.e., high reward sensitivity). This interpretation is consistent with the model of perfectionism proposed by Hewitt and Flett, in which self-oriented perfectionism, or the constant pursuit to be flawless, interacts with negative life events to lead to depression (Blatt 1995). That risk taking in individuals with higher levels of depressive symptoms depended on whether the preceding trial was rewarded or punished may reflect this interaction between high perfectionism and a sense of failure. This speculation, however, should be empirically examined in

future research. Further, it might be explored more in the context of the stress generation hypothesis, considering the hypothesis's emphasis on certain personality factors (e.g., perfectionism) contributing to depression's generation and maintenance of stress (Safford et al. 2008).

The current sample had higher levels of depressive symptoms than would be expected of a college sample. Specifically, our mean (21.99) is higher than the traditionally recommended cut-off score of 16. However, this cut-off score overestimates depression in a nonclinical sample, with 45% of college students being classified as severely depressed (Santor et al. 1995). Relatedly, Santor et al. suggested that a CES-D score of 34 would be a more appropriate cut-off score that would estimate around 8% as severely depressed in college students, a prevalence rate more consistently found in this population using other depression measures such as Beck Depression Inventory-II (e.g., Beiter et al. 2015; Santor et al. 1995). Indeed, even the participants whose CES-D scores were at 1 SD above our sample mean endorsed a mild level of depressive symptoms according to Olinio et al. (2013). Thus, caution is warranted in using and interpreting the cut-off score. We do, however, acknowledge that the reliability of the CES-D scores in the current sample, although acceptable, was lower than in previous studies, which might be partly due to the fact that the participants were asked about depressive symptoms over the past 2 months.

We did not specifically recruit participants with major depressive disorder, and a clinical sample with greater impairments in functioning may respond differently to the task. Thus, future studies should examine risk taking in a clinical sample and also investigate the influence of perfectionism on risk taking in subclinical and clinical depression. Our sample also consisted of predominantly female participants; whether the findings extend to individuals who identify by a different gender should be investigated. The study should also be replicated with a more racially diverse sample, considering that our sample was predominantly white.

To note, the current finding was based on cross-sectional data, and, thus, we cannot make any causal conclusion. The relation between depressive symptoms and punishment sensitivity should also be examined using more naturalistic methods (e.g., diary study). Further, the current design did not permit investigating the difference in anticipatory and consummatory aspects of reward processing (Berridge and Robinson 2003) in risk taking, which should be investigated in future studies. Additionally, it would be important to explore the persistence of heightened punishment sensitivity and how this relates to depressive symptoms, which could help elucidate a critical factor maintaining depression.

We hypothesized that the tasks used to measure risk taking may have contributed to the inconsistent findings regarding punishment sensitivity in depression. However, there are other possible factors. Depression is highly comorbid

with anxiety (Kessler et al. 2008), which is associated with greater punishment sensitivity (Giorgetta et al. 2012). Further, different combinations of depressive symptoms may contribute to different findings on punishment sensitivity. Specifically, anhedonia may be more relevant in the association between depressive symptoms and reward sensitivity. Therefore, future studies should examine the specific symptoms of depression, particularly anhedonia, and comorbid anxiety symptoms in their relation to reward versus punishment sensitivity.

Notwithstanding these limitations, we found that individuals with higher levels of depressive symptoms pumped fewer times after a punishment trial than a reward trial, indicating heightened punishment sensitivity. When faced with setbacks, individuals with high levels of depressive symptoms may deem negative outcomes more likely and avoid risk-taking, even with potentially high long-term rewards. Avoiding risks can minimize opportunities to experience rewards that challenge negative schemas, thereby maintaining or exacerbating depressive symptoms.

Author contributions All authors contributed to the study conception and design. Material preparation and data collection were performed by Amanda Kutz and K. Lira Yoon. The first draft of the manuscript was written by Dahyeon Kim and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding No funds, grants, or other support was received.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Bechara, A., Damasio, A., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7–15. [https://doi.org/10.1016/0010-0277\(94\)90018-3](https://doi.org/10.1016/0010-0277(94)90018-3).
- Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *American Journal of Psychiatry*, *165*(8), 969–977. <https://doi.org/10.1176/appi.ajp.2008.08050721>.
- Beiter, R., Nash, R., McCrady, M., Rhoades, D., Linscomb, M., Clarahan, M., & Sammut, S. (2015). The prevalence and correlates of depression, anxiety, and stress in a sample of college students. *Journal of Affective Disorders*, *173*(1), 90–96. <https://doi.org/10.1016/j.jad.2014.10.054>.
- Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences*, *26*(9), 507–513. [https://doi.org/10.1016/S0166-2236\(03\)00233-9](https://doi.org/10.1016/S0166-2236(03)00233-9).
- Blatt, S. J. (1995). The destructiveness of perfectionism: Implications for the treatment of depression. *American Psychologist*, *50*(12), 1003–1020. <https://doi.org/10.1037/0003-066X.50.12.1003>.
- Buelow, M. T., & Blaine, A. L. (2015). The assessment of risky decision making: A factor analysis of performance on the Iowa Gambling Task, Balloon Analogue Risk Task, and Columbia Card Task. *Psychological Assessment*, *27*(3), 777–785. <https://doi.org/10.1037/a0038622>.
- Cavanagh, J. F., Bismark, A., Frank, M. J., & Allen, J. J. (2011). Larger error signals in major depression are associated with better avoidance learning. *Frontiers in Psychology*, *2*(331), 24–29. <https://doi.org/10.3389/fpsyg.2011.00331>.
- Carvalho, J. P., & Hopko, D. R. (2011). Behavioral theory of depression: Reinforcement as a mediating variable between avoidance and depression. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*(2), 154–162. <https://doi.org/10.1016/j.jbtep.2010.10.001>.
- Enns, M. W., Cox, B. J., & Borger, S. C. (2001). Correlates of analogue and clinical depression: A further test of the phenomenological continuity hypothesis. *Journal of Affective Disorders*, *66*(2–3), 175–183. [https://doi.org/10.1016/S0165-0327\(00\)00305-0](https://doi.org/10.1016/S0165-0327(00)00305-0).
- Deisenhammer, E. A., Schmid, S. K., Kemmler, G., Moser, B., & Delazer, M. (2018). Decision making under risk and under ambiguity in depressed suicide attempters, depressed non-attempters and healthy controls. *Journal of Affective Disorders*, *226*, 261–266. <https://doi.org/10.1016/j.jad.2017.10.012>.
- Giorgetta, C., Grecucci, A., Zuanon, S., Perini, L., Balestrieri, M., Bonini, N., et al. (2012). Reduced risk-taking behavior as a trait feature of anxiety. *Emotion*, *12*(6), 1373–1383. <https://doi.org/10.1037/a0029119>.
- Gorlyn, A., Keilp, J. G., Oquendo, M. A., Burke, A. K., & Mann, J. J. (2013). Iowa Gambling Task performance in currently depressed suicide attempters. *Psychiatry Research*, *207*(3), 150–157. <https://doi.org/10.1016/j.psychres.2013.01.030>.
- Henriques, J. B., & Davidson, R. J. (2000). Decreased responsiveness to reward in depression. *Cognition & Emotion*, *14*(5), 711–724. <https://doi.org/10.1080/02699930050117684>.
- Hevey, D., Thomas, K., Laureano-Schelten, S., Looney, K., & Booth, R. (2017). Clinical depression and punishment sensitivity on the BART. *Frontiers in Psychology*, *8*(670), 1–70. <https://doi.org/10.3389/fpsyg.2017.00670>.
- Hewitt, P. L., & Flett, G. L. (1991). Dimensions of perfectionism in unipolar depression. *Journal of Abnormal Psychology*, *100*(1), 98–101. <https://doi.org/10.1037/0021-843X.100.1.98>.
- Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*(1), 101–117. <https://doi.org/10.3758/BF03192848>.
- Holmes, A. J., & Pizzagalli, D. A. (2008). Spatiotemporal dynamics of error processing dysfunctions in major depressive disorder. *Archives of General Psychiatry*, *65*(2), 179–188. <https://doi.org/10.1001/archgenpsychiatry.2007.19>.
- Kessler, R. C., Gruber, M., Hettema, J. M., Hwang, I., Sampson, N., & Yonkers, K. A. (2008). Comorbid major depression and generalized anxiety disorders in the National Comorbidity Survey follow-up. *Psychological Medicine*, *38*(3), 365–374. <https://doi.org/10.1017/S0033291707002012>.
- Leahy, R. L., Tirch, D. D., & Melwani, P. S. (2012). Processes underlying depression: Risk aversion, emotional schemas, and psychological flexibility. *International Journal of Cognitive Therapy*, *5*(4), 362–379. <https://doi.org/10.1521/ijct.2012.5.4.362>.
- Lejuez, C. W., Aklin, W. M., Jones, H. A., Richards, J. B., Strong, D. R., Kahler, C. W., & Read, J. P. (2003). The Balloon Analogue Risk Task (BART) differentiates smokers and nonsmokers. *Experimental and Clinical Psychopharmacology*, *11*, 26–33. <https://doi.org/10.1037/1064-1297.11.1.26>.
- Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., et al. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART).

- Journal of Experimental Psychology Applied*, 8, 75–84. <https://doi.org/10.1037/1076-898X.8.2.75>.
- Leykin, Y., Dunn, L. B., & Munoz, R. F. (2018). The effect of depression on the decision to join a clinical trial. *Journal of Consulting and Clinical Psychology*, 85(7), 751–756. <https://doi.org/10.1037/ccp0000212>.
- Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, 113(1), 181–190. <https://doi.org/10.1037/0033-2909.113.1.181>.
- Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods*, 222, 250–259.
- Olino, T. M., Yu, L., McMakin, D. L., Forbes, E. E., Seeley, J. R., Lewinsohn, P. M., & Pilkonis, P. A. (2013). Comparisons across depression assessment instruments in adolescence and young adulthood: An item response theory study using two linking methods. *Journal of Abnormal Child Psychology*, 41(8), 1267–1277. <https://doi.org/10.1007/s10802-013-9756-6>.
- Pizzagalli, D. A., Iosifescu, D., Hallett, L. A., Ratner, K. G., & Fava, M. (2008). Reduced hedonic capacity in major depressive disorder: Evidence from a probabilistic reward task. *Journal of Psychiatric Research*, 43(1), 76–87. <https://doi.org/10.1016/j.jpsyc hires.2008.03.001>.
- Pizzagalli, D. A., Jahn, A. L., & O’Shea, J. P. (2005). Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. *Biological Psychiatry*, 57(4), 319–327. <https://doi.org/10.1016/j.biopsych.2004.11.026>.
- Pleskac, T. J., Wallsten, T. S., Wang, P., & Lejuez, C. W. (2008). Development of an automatic response mode to improve the clinical utility of sequential risk-taking tasks. *Experimental and Clinical Psychopharmacology*, 16(6), 555–564. <https://doi.org/10.1037/a0014245>.
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31, 437–448. <https://doi.org/10.3102/10769986031004437>.
- Radloff, L. (1977). The CES-D scale a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385–401.
- Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: A systematic review and meta-analysis. *Psychological Medicine*, 44(10), 2029–2040. <https://doi.org/10.1017/S0033291713002535>.
- Roepke, A. M., & Seligman, M. E. P. (2016). Depression and prospection. *British Journal of Clinical Psychology*, 55, 23–48. <https://doi.org/10.1111/bjc.12087>.
- Safford, S. M., Alloy, L. B., Abramson, L. Y., & Crossfield, A. G. (2008). Negative cognitive style as a predictor of negative life events in depression-prone individuals: A test of the stress generation hypothesis. *Journal of Affective Disorder*, 99(1–3), 147–154. <https://doi.org/10.1016/j.jad.2006.09.003>.
- Santor, D. A., Zuroff, D. C., Ramsay, J. O., Cervantes, P., & Palacios, J. (1995). Examining scale discriminability in the BDI and CES-D as a function of depressive severity. *Psychological Assessment*, 7(2), 131–139. <https://doi.org/10.1037/1040-3590.7.2.131>.
- Shean, G., & Baldwin, G. (2008). Sensitivity and specificity of depression questionnaires in a college-age sample. *The Journal of Genetic Psychology*, 169(3), 281–292. <https://doi.org/10.3200/GNTP.169.3.281-292>.
- Smoski, M. J., Lynch, T. R., Rosenthal, M. Z., Cheavens, J. S., Chapman, A. L., & Krishnan, R. R. (2008). Decision-making and risk aversion among depressive adults. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(4), 567–576. <https://doi.org/10.1016/j.jbtep.2008.01.004>.
- Steffens, D. C., Wagner, H. R., Levy, R. M., Horn, K. A., & Krishnan, K. R. R. (2001). Performance feedback deficit in geriatric depression. *Biological Psychiatry*, 50(5), 358–363. [https://doi.org/10.1016/S0006-3223\(01\)01165-9](https://doi.org/10.1016/S0006-3223(01)01165-9).

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.