

Social Psychology

Sadder ≠ Wiser: Depressive Realism Is Not Robust to Replication

Amelia S. Dev¹ , Don A. Moore² ^a, Sheri L. Johnson³, Karin T. Garrett⁴

¹ Psychology, University of California, Berkeley, US; University of Miami, US; ² Haas School of Business, University of California Berkeley, Berkeley, US,

³ Psychology, University of California Berkeley, US, ⁴ Haas School of Business, University of California Berkeley, US

Keywords: depression, depressive realism, overconfidence

<https://doi.org/10.1525/collabra.38529>

Collabra: Psychology

Vol. 8, Issue 1, 2022

The theory of depressive realism holds that depressed individuals are less prone to optimistic bias, and are thus more realistic, in assessing their control or performance. Since the theory was proposed 40 years ago, many innovations have been validated for testing cognitive accuracy, including improved measures of bias in perceived control and performance. We incorporate several of those innovations in a well-powered, pre-registered study designed to identify depressive realism. Amazon MTurk workers ($N = 246$) and undergraduate students ($N = 134$) completed a classic contingency task, an overconfidence task, and measures of mental health constructs, including depression and anxiety. We measured perceived control throughout the contingency task, allowing us to compare control estimates at the trial-level to estimates assessed at task conclusion. We found no evidence that depressive symptoms relate to illusory control or to overconfidence. Our results suggest that despite its popular acceptance, depressive realism is not replicable.

Introduction

Depressive realism, the idea that depression is associated with more accurate perceptions of personal control, is widely accepted. Alloy and Abramson's (1979) original paper, "Judgment of contingency in depressed and nondepressed students: Sadder but wiser?" has been cited over two thousand times. Depressive realism has been a focus of more than 75 empirical studies (M. T. Moore & Fresco, 2012), five books (Alloy et al., 1990; Alloy & Abramson, 1988; Baker et al., 2012; Feltham, 2017; Murphy et al., 2005), and articles published in *The Guardian* (Burkeman, 2006), *Psychology Today* (Burton, 2012), *The New Yorker* (Konnikova, 2014), and *Vice* (Bucklin, 2017). Taylor and Brown (1988) built their influential theory of "positive illusions" on these claims; their 1988 article has been cited over ten thousand times. Wong et al. summarize the scholarly consensus: "The depressed exhibit a remarkable degree of realism in judgments about their personal and social worlds, whereas the less depressed tend to exhibit unrealistic optimism" (2006, p. 284).

Despite its pervasive influence, there is mixed empirical support for depressive realism. Some have failed to replicate the effect (for reviews see Ackermann & DeRubeis, 1991; Allan et al., 2007; and M. T. Moore & Fresco, 2012), even in clinical samples (Shruti Venkatesh et al., 2018). A meta-analysis of 75 studies indicated substantial variability in effect sizes ($d = -.09$ to $.14$) (M. T. Moore & Fresco, 2012),

indicating the need to refine our understanding of this effect.

In contrast to depressive realism, other research suggests that individuals with depression are prone to cognitive bias. Studies have found that depressed individuals underestimate their true capabilities (Fu et al., 2005, 2012; Stone et al., 2001), experience overly negative self-focused thoughts (Gotlib & Joormann, 2010), forecast excessively negative moods (Zetsche et al., 2019), and generally show biases as pronounced as their nondepressed counterparts (Dobson & Pusch, 1995), if not more so (Dunning & Story, 1991). These negative biases have been found to precede the onset of depression (Everaert et al., 2012; Lewinsohn et al., 2001; Mathews & MacLeod, 2005; Rude et al., 2003; Segal et al., 2006).

Despite abundant evidence establishing the presence of biases in depression *and* numerous failures to replicate the depressive realism effect, research on depressive realism continues to be published and cited, indicating that psychological scientists and laypeople retain some belief in its plausibility. In this paper, we seek to establish if the sadder are indeed wiser. We set out to test the depressive realism effect; to do so, we conducted a direct replication test of Alloy and Abramson (1979) but also implemented several design improvements, incorporating additional, validated measures of bias and considering several clinical features that could inform findings. Our study design benefits from methodological innovations since the original studies,

a Address correspondence to: dm@berkeley.edu

including design elements and boundary conditions to help identify the depressive realism effect.

In Alloy and Abramson's original depressive realism study (1979), participants estimated the control their responses (button pushes) gave them over an outcome (a lightbulb turning on). The outcome measure used in this original study, though, did not compare subjective control estimates to actual control, thereby making it impossible to measure bias. In one meta-analysis, only 36 of 75 depressive realism studies employed measures of bias. These 36 studies found smaller effects (weighted mean $d = -.03$) than did studies without an objective standard (weighted mean $d = -.15$; M. T. Moore & Fresco, 2012). Furthermore, only 15 of these 36 studies evaluated depressive realism via a contingency task. We employed a contingency task that allowed for objective assessment of the degree to which participants over- or underestimate their control. We also aimed to test alternate hypotheses of the depressive realism effect. First, we test the response criterion hypothesis, which posits that depressive ideation will influence post-task assessments of overall control more strongly than trial-level estimates (Allan et al., 2007). The logic behind this hypothesis grows from depression's memorial and attentional biases, which operate more strongly on recollections and high-level construals than on immediate and direct assessments (Marchetti et al., 2018). To test this hypothesis, we administer measures of perceived control throughout the task and also at the end. Second, given that motor response rates profoundly influence beliefs about control and are often diminished among those with depression (Blanco et al., 2012), we measure whether the frequency of responding helps explain how depression relates to beliefs about control (Blanco et al., 2009). Finally, we ask participants to estimate outcome probabilities based on both possible responses (pressing and not pressing the button), since true contingencies exist for both.

Moore and Fresco (2012) conclude that future research should seek to clarify if there are any situations which *do* yield a robust depressive realism effect, and their suggestions guided some of our additional design decisions. Specifically, we include a high-control condition that allows us to test for accuracy when accuracy demands high estimates of control. In addition, we consider the generalizability of depressive realism to an additional overconfidence task. The latter is a validated laboratory task that assesses the accuracy of self-assessment. Just as we measure illusory control in the contingency task against objective reality, we compare self-assessments with the truth in the overconfidence task. Finding increased accuracy among individuals with higher depressive symptoms in the overconfidence task would support the theory of depressive realism.

We also consider key issues in symptom assessment. Because taxometric analyses suggest that depression is best conceptualized as a dimensional rather than categorical construct (Prisciandaro & Roberts, 2005), and dichotomizing participants as 'depressed' or 'not depressed' prevents analyzing symptoms continuously (Soderstrom et al., 2011), we examine depressive symptoms over a range of severity. Moore and Fresco's (2012) meta-analysis suggests

that on average, both clinically depressed and subsyndromal individuals show *overestimation* of control, but that subsyndromal participants demonstrate the strongest depressive realism effect. Thus, we focus on subsyndromal participants, who may be particularly likely to show the effect. Previous studies have also found more robust effects for self-ratings interview-based ratings of symptoms; accordingly, we focus on self-ratings here (M. T. Moore & Fresco, 2012). The original depressive realism study (Alloy & Abramson, 1979) examined only current depression; to replicate findings that biases improve after recovery (S. Venkatesh, 2013), we examine lifetime as well as current depression symptom severity. We also measure clinical characteristics that may moderate the depressive realism effect, such as negative affective state, hypomanic tendencies, trauma history, and anxiety (Ackermann & DeRubeis, 1991; Stern & Berrenberg, 1979; Wong et al., 2006).

Method

Transparency and Openness

We report how we determined our sample size, all measures, experimental conditions, and data exclusions. Data were analyzed using R. We preregistered the study design and primary analyses. Our materials, including survey instruments, data, analysis code, and copies of the preregistration documents, are posted on the Open Science Framework [here](#).

Participants

Our study includes two samples. Our first sample, collected in Fall 2020, came from Amazon Mechanical Turk. We sought to generalize and replicate our results with a second sample of undergraduates at a large, public U.S. university in Spring 2021.

Sample One

We based our sample size on a priori power analyses of three effects reported in experiments 2-4 of the original Alloy and Abramson (1979) paper. Averaging across these three effect sizes, an a priori power analysis (conducted in G*Power) suggested an average sample size of 40 to obtain 90% power for detecting the effect. To maximize our chances of detecting a true effect, we followed suggestions to conduct conservative power analyses based on the lower bound of the 90% confidence interval around each effect (Anderson & Maxwell, 2017). Results of these more conservative power analyses suggested $N = 252$ for 90% power to detect a depressive realism effect (see Supplement for more information).

Sample One (S1) included participants age 18 and older, located in the United States, recruited through Amazon Mechanical Turk (MTurk). To obtain at least 25 participants each with minimal, mild, moderate or severe depressive symptoms, we pre-screened 800 participants with the Inventory to Diagnose Depression - Current (IDD-C; cut-offs based on Rogers et al. (2005) and invited 339 to take the full survey (180 accepted the survey). Another 153 partic-

Participants completed the survey without prescreening. After 57 exclusions for missing data, 29 exclusions for failing a comprehension check, and 1 exclusion of a duplicated response, Sample One included 246 participants (53% male, 1% gender nonbinary) with complete data. Of these, 128 had minimal depressive symptoms (IDD-C ≤ 15), 38 had mild depressive symptoms (IDD-C 16-24), 45 had moderate depressive symptoms (IDD-C 25-41), and 35 had severe depressive symptoms (IDD-C > 42).

Sample Two

Sample Two consisted of undergraduates from a large, public U.S. university, who participated in the study in return for course credit. While we sought to prescreen participants to oversample with higher depression symptom severity, we were constrained by the sampling pool. After 16 exclusions for missing data, 24 exclusions for failing a comprehension check, and 9 exclusions of duplicate responses, Sample Two included 136 participants (55.9% male, 1.5% gender nonbinary) with complete data. Of these, 14 had minimal depressive symptoms (IDD-C < 15), 104 had mild depressive symptoms (IDD-C 16-24), 14 had moderate depressive symptoms (IDD-C 25-41), and 2 had severe depressive symptoms (IDD-C > 42).

Procedure

Participants completed informed consent, the contingency task, the overconfidence task, and then questionnaires online using the Qualtrics survey platform.

Contingency Task

The contingency task was a revised version of the Alloy and Abramson (1979) task. In each of 40 trials, participants chose whether to press a button, after which a lightbulb or a black screen appeared. The survey randomly assigned participants to one of three experimental conditions. In two “zero contingency” conditions, there was no relationship between their action (the button) and the outcome (the lightbulb). Regardless of whether they pressed the button, the lightbulb appeared with either 25% or 75% chance (in the first and second zero contingency conditions, respectively). In the third “positive contingency” condition, the lightbulb appeared with a 75% chance following button press and never otherwise.

Participants read that “On every round of the lightbulb task, you will have the chance to press or not press a virtual button. You will then either see a yellow lightbulb icon or a black box. Your job is to figure out how your button responses (pressing or not pressing) affect the probability of the yellow light icon appearing.” To replicate previous work, after the final trial, the survey asked “How much control did you have over the lightbulb?” on a 0 (*none*) to 100 (*complete control*) scale. To improve upon this single, point estimate, we also used a more precise elicitation of subjective control. Every ten trials, participants estimated the likelihood (on a 0 to 100% scale) of the bulb lighting if they did press the button on the next trial, and if they did not.

The survey randomly asked half of participants about the outcome of the lightbulb appearing and half about the outcome of the lightbulb *not* appearing. Participants read that more accurate predictions increased their chances of winning a drawing for a \$50 prize.

We calculated three key measures. First, frequency of responding is the percent of trials on which participants pressed the button. The second and third measures were control bias and outcome prediction bias: both measured absolute values of subjective control minus actual control. Control bias, measured at the end of the task, was the difference between final ratings of control (“How much control did you have over the lightbulb?”) minus actual control (zero in the zero contingency conditions, 75 in the positive contingency condition). Outcome prediction bias, measured during the task, was the absolute difference between actual and estimated control. Estimated control was the reported probability that the lightbulb would light after (1) pressing and (2) not pressing the button, averaged across the estimates made every ten trials.

Overconfidence Task

Participants completed ten items from the Raven’s Progressive Matrices (RPM; Raven, 1963). They then provided two subjective probability distributions for performance on the Raven’s task: they estimated probabilities that they obtained each of the eleven possible scores (from zero correct to ten correct), for the self and for a random other (participants read: “We randomly selected one other participant out of the large number of people who have also completed this survey.”). As is standard for overconfidence tasks (D. A. Moore & Healy, 2008), we calculated three measures (see Supplement for detail). First, overestimation compares participants’ estimated scores for themselves to their actual scores. Second, overplacement compares participants’ predictions for how much better they score than others, and adjusts for how much better they actually do than others. Third, overprecision captures excessive certainty in participants’ estimates of others’ scores, assessed via the Gini index (a measure of concentration in the distribution of score estimates) and item confidence (maximum confidence attached to any single score when estimating the probability of different scores for a random other).

Questionnaires

The final section of the survey assessed the valence of participants’ current mood, from 0 (*not negative at all*) to 100 (*very negative*), and included several validated, psychometrically sound self-report questionnaires in randomized order: the Negative generalization subscale from the Attitudes Toward Self-Revised (ATS-R; Carver et al., 1988), the Hypomanic Personality Scale (HPS; Eckblad & Chapman, 1986), the Inventory to Diagnose Depression-Lifetime (IDD-L; Zimmerman & Coryell, 1987) the Mood and Anxiety Symptom Questionnaire-D30 (MASQ-D30; Wardenaar et al., 2010), and the Risky Families Questionnaire (RF; Taylor et al., 2004). Participants also completed the IDD-C, either in this final section of the survey or as part of pre-

Table 1. Descriptive Statistics for End-of-task Subjective Control.

Condition	Actual control	Sample One			Sample Two		
		Subjective control M (SD)	t^a	df	Subjective control M (SD)	t	df
Zero contingency, 25% probability	0	27.64 (29.33)	8.59***	82	18.15 (15.21)	7.55***	39
Zero contingency, 75% probability	0	34.23 (30.54)	9.84***	76	36.83 (26.51)	9.01***	41
75% positive contingency	75	61.54 (28.53)	-4.22***	79	65.02 (26.97)	-2.51*	45

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

^a Results from one-sample t -tests comparing the end-of-task subjective control measure to actual control.

screening. Sample Two included additional measures: the Sheehan Disability Scale (SDS; Sheehan et al., 1996), current and lifetime anxiety diagnosis and treatment, as well as current and lifetime depression diagnosis and treatment.

Analysis Plan

Our preregistered analysis plan used an ANOVA to test the results of our direct replication, and parallel hierarchical linear models to examine the effect of continuous IDD-C scores on our overconfidence and contingency task measures.

Results

All analyses were conducted using R statistical software. Skew and kurtosis (≤ 1.47), and Cronbach's alpha values ($\geq .79$) were within acceptable ranges for all self-report scales (see Table S1 in the Supplement).

Contingency Task Results

Table 1 shows descriptive statistics for the end-of-task subjective control measure (i.e., participants' answer when asked "How much control did you have over the light-bulb?"). It also presents one-sample t -tests comparing the end-of-task subjective control measure to actual control.

In conditions with zero true contingency, end-of-task subjective control was significantly higher than actual control, indicating that participants overestimated control. In the positive contingency condition, end-of-task subjective control was significantly lower than actual control, indicating that participants underestimated control. Underestimation of control when control is high is a necessary consequence of regressive estimates of control given uncertainty and noise in participants' control estimates (Gino et al., 2011).

Direct Replication

To test the depressive realism hypothesis, we conducted a 4 (IDD-C severity) \times 3 (contingency condition) \times 2 (gender) ANOVA estimating control bias.¹ Depressive realism predicted a main effect of IDD-C.

In Sample One, the main effect of IDD-C score was indeed significant, $F(3, 215) = 3.88, p = .01, \eta p^2 = 0.05$, as was the main effect of contingency condition ($F(2, 215) = 69.06, p < .001, \eta p^2 = 0.39$). However, as Figure 1 shows, this result contradicts the sadder but wiser hypothesis, since more severe depression is associated with greater overestimates of control. We did not find any significant effect of the IDD-C in Sample Two ($F(3, 110) = 0.821, p = 0.49$). Our attempt to replicate the interaction Alloy and Abrahamson report between Contingency condition and IDD-C score was not successful in Sample One, ($F(6, 215) = 1.09, p = .37, \eta p^2 = 0.03$), or in Sample Two ($F(4, 110) = 1.23, p = 0.30, \eta p^2 = .04$).

The main effect of contingency condition was significant in Sample One ($F(2, 215) = 69.06, p < .001, \eta p^2 = 0.39$) and in Sample Two ($F(2, 110) = 41.597, p < .001, \eta p^2 = 0.43$). This effect is a consequence of the fact that participants' overestimates of control are limited to the zero contingency conditions in which they lacked control.

Hierarchical linear regressions. To examine IDD-C scores as continuous variables, we constructed three parallel hierarchical linear regressions estimating control bias, outcome prediction bias, and frequency of responding. Table 2 shows the preregistered models.² Steps 1 and 2 included variables to replicate Alloy and Abramson (1979) and Steps 3-5 considered variables that may influence this effect. We only report significant variables below, but full model results appear in the supplementary material. We also conducted supplemental analyses: we tested models including polynomial terms to probe for potential curvilinear effects, and tested models including a trial-level random effect to better account for measurement error in the

1 Because Alloy and Abramson (1979) reported finding a depressive realism effect among only female participants, our preregistered analysis includes gender. We do not find any significant main or interaction effects of gender.

2 We erroneously omitted IDD-C score \times contingency type from the pre-registration, but it is essential to understanding whether true contingency influences any potential depressive realism effect. Likewise, we realized after pre-registering that we needed a model including frequency of responding to test for a depressive realism effect in control and outcome bias.

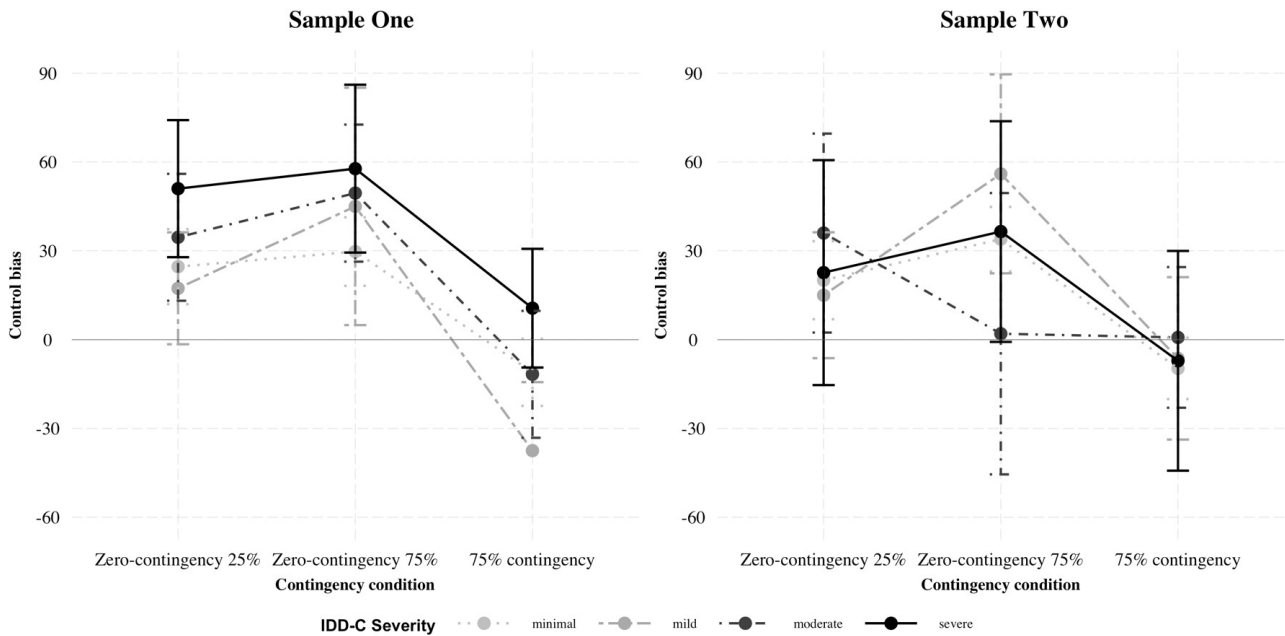


Figure 1. Interaction Effect Between Contingency Condition and IDD-C Score on Control Bias

contingency task. Results from these supplemental analyses did not reveal quantitative or qualitative differences from our main analyses (see Supplement for full results). Note that supplemental analyses were not preregistered.

Control bias. Step 1 accounted for significant variance in both samples (S1: $F(3, 236) = 46.4, p < .001, R^2 = 0.36$; S2: $F(3, 124) = 29.01, p < .00001, R^2 = 0.40$). No other model steps accounted for significantly increased variance (S1: $R^2_{\text{change}} < .025, p > .06$; S2: $R^2_{\text{change}} < .023, p > .294$). Contingency type (positive vs. 0) was associated with control bias (S1: $\beta = -0.65, t(236) = -10.70, p < .001$; S2: $\beta = -0.74, t(124) = -9.242, p < .001$). Higher IDD-C scores predicted higher control bias in Sample One ($\beta = 0.18, t(236) = 3.38, p < .001$); however, this result does not replicate in Sample Two ($\beta = 0.05, t(124) = 0.723, p = 0.47$). Figure 2 shows these results.

Given the significant effects for IDD-C in Sample One, we conducted preregistered analyses to examine depression and anxiety symptoms separately, by replacing the IDD-C with the MASQ Anhedonic Depression and Anxious Arousal subscales. When compared to Step 1, the Step 2 model predicted significantly increased variance, $F(8, 232) = 23.75, p < .001, R^2 = 0.43$. As above, contingency type (positive vs. zero) significantly affected control bias, $\beta = -0.63, t(231) = -3.01, p = .003$. Anhedonia Depression did not significantly relate to control bias, but greater Anxious Arousal accompanied increased control bias, $\beta = 0.26, t(231) = 3.72, p < .001$. The interaction between contingency type and anxious arousal was also significant, $\beta = -0.15, t(231) = -2.55, p = .01$. No other model steps accounted for significantly increased variance, all $R^2_{\text{change}} < .009, p > .46$. See Figure 3. It did not make sense to conduct these post-hoc analyses in Sample Two given the null effects of IDD-C on control bias.

Outcome prediction bias. As noted above, outcome prediction bias was calculated as the absolute difference between actual control and estimated control, wherein es-

timated control was the reported probability that the lightbulb would light after (1) pressing and (2) not pressing the button, averaged across the estimates participants made every ten trials. Table 3 shows descriptive statistics for estimated control and results of one-sample t-tests comparing estimated control to actual control.

Mirroring results for the end-of-task subjective control measure, participants' estimated control was significantly higher than actual control in the zero contingency conditions and significantly lower than actual control in the positive contingency condition.

Overall, results for outcome prediction bias were consistent across samples. The Step 1 model accounted for significant variance in both samples (S1: $F(3, 236) = 142.7, p < .001, R^2 = 0.64$; S2: $F(3, 124) = 50.3, p < .001, R^2 = 0.54$). In this model, contingency type (positive vs. zero contingency) was significant (S1: $\beta = -0.84, t(236) = -18.45, p < .001$; S2: $\beta = -0.83, t(124) = -11.767, p < .001$), as was contingency probability (75% vs. 25%) (S1: $\beta = 0.09, t(236) = 2.06, p = .04$; S2: $\beta = 0.21, t(124) = 2.97, p < .001$). No other model steps accounted for significantly increased variance, (S1: $R^2_{\text{change}} < .004, p > .36$; S2: $R^2_{\text{change}} < .03, p > .403$). As with control bias, Figure 4 illustrates that participants overestimate control in the zero contingency conditions and underestimate it in the positive contingency condition. IDD-C score was not significantly associated with outcome prediction bias in either sample. Moreover, our results allow us to track the development of beliefs over time. As detailed in the Supplement, those results are consistent with those reported above: participants update their beliefs in the direction of the control they experience, and IDD-C scores do not affect either learning or reported control.

Frequency of responding. Participants pressed the button on roughly half of all trials (S1: $M = 57\%, SD = 19\%$,

Table 2. Hierarchical Regression Models for the Contingency Task Measures

Step	Variable(s)
Step 1	Contingency type (zero contingency = 0 and positive contingency = 1) Contingency outcome probability (25% or 75%) IDD-C score
Step 2	IDD-C score x contingency type IDD-C score x contingency outcome probability
Step 3	Negative affect IDD-L score
Step 4 ^a	Negative affect x IDD-L score Treatment history (dummy coded: 0 (no) and 1 (yes)) Hypomanic Personality Scale (HPS) score Risky Families (RF) score Attitudes Toward Self (ATS) - Negative Generalization score Sheehan Disability Scale (SDS) score Depression diagnosis Depression treatment Anxiety diagnosis Anxiety treatment
Step 5 ^b	Frequency of responding

Note. IDD-C = Inventory to Diagnose Depression – Current. IDD-L = Inventory to Diagnose Depression – Lifetime.

^a Additional variables included in Sample Two were depression diagnosis, depression treatment, anxiety diagnosis, anxiety treatment, as well as the Sheehan Disability Scale. Diagnosis and treatment variables captured both current diagnosis/treatment, lifetime diagnosis/treatment, or no history of diagnosis/treatment.

^b Step 5 was included for the control and outcome prediction bias models only

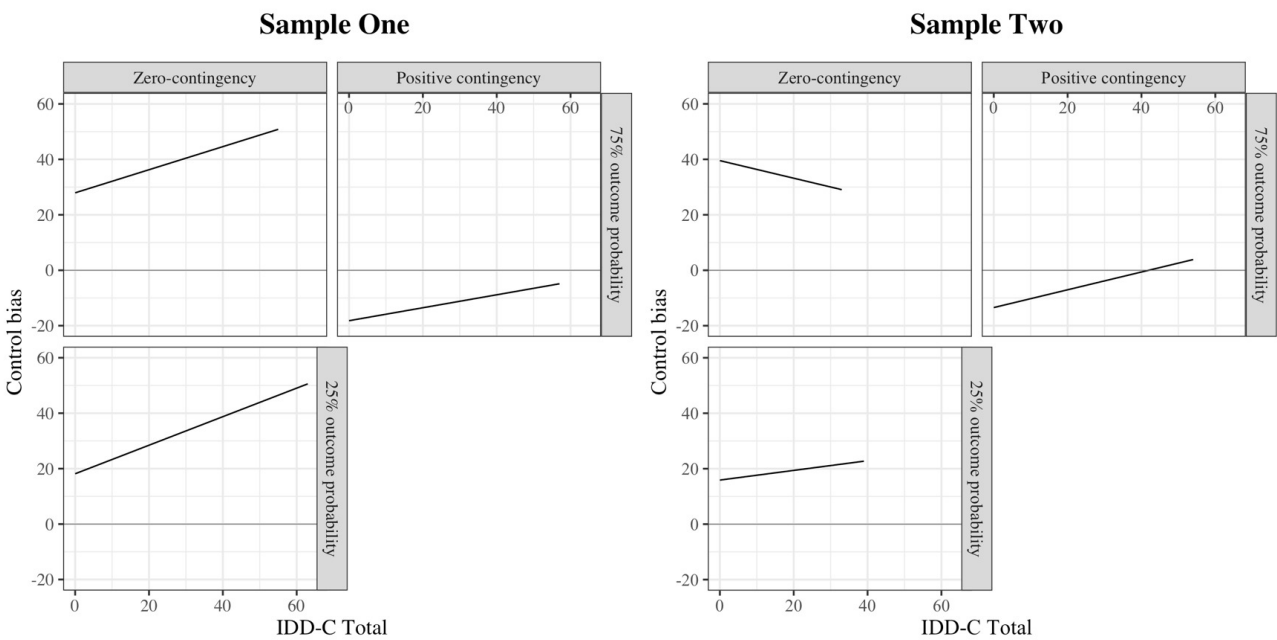


Figure 2. The Effect of IDD-C Score on Control Bias Across Contingency Conditions.

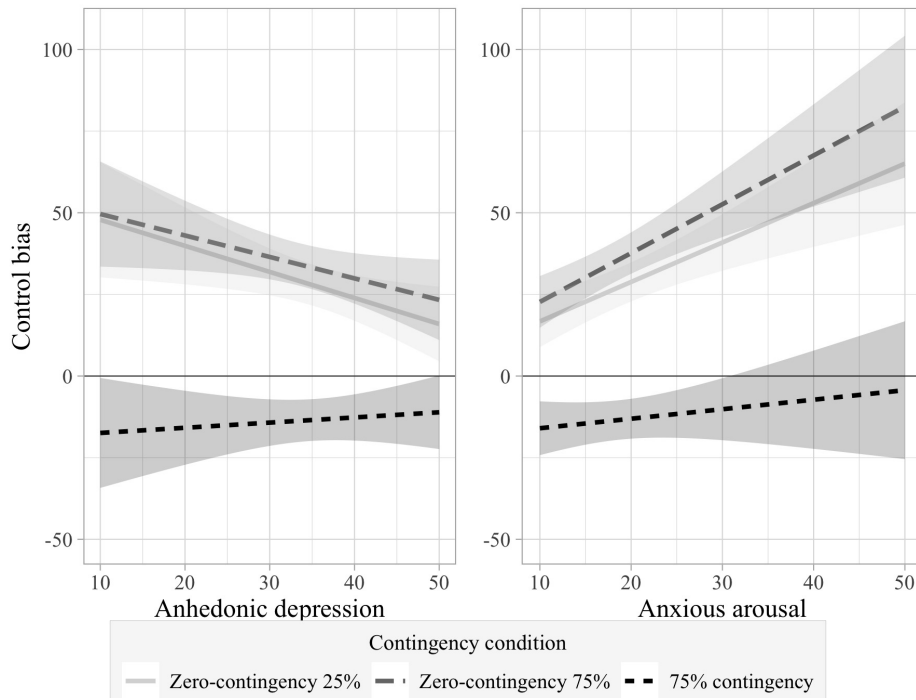


Figure 3. Effects of MASQ Anhedonic Depression and Anxious Arousal on Control Bias

Note. Shaded areas depict 95% confidence intervals.

Table 3. Descriptive Statistics for Estimated Control.

Condition	Actual control	Sample One			Sample Two		
		Estimated control M (SD)	t	df	Subjective control M (SD)	t	df
Zero contingency, 25% probability	0	11.33 (11.16)	9.25***	82	10.83 (11.7)	5.85***	39
Zero contingency, 75% probability	0	17.51 (16.35)	9.40***	76	23.71 (20.08)	7.65***	41
75% positive contingency	75	38.2 (24.6)	-13.4***	79	49.18 (24.15)	-7.25***	45

Note. * $p < .05$ ** $p < .01$ *** $p < .001$

^a Results from one-sample t -tests comparing the end-of-task subjective control measure to actual control.

Range = 3% – 100%; S2: $M = 58\%$, $SD = 17\%$, Range = 28% – 100%).

In Sample One and Sample Two, the Step 1 model accounted for significant variance (S1: $F(3, 236) = 4.76$, $p = .003$, $R^2 = 0.04$; S2: $F(3, 124) = 7.88$, $p < .001$, $R^2 = 0.14$). Contingency type was significantly associated with frequency of responding in both samples (S1: $\beta = 0.17$, $t(236) = 2.29$, $p = 0.02$; S2: $\beta = 0.26$, $t(124) = 2.68$, $p = 0.008$). The IDD-C score was only marginally significant in Sample Two, with higher depression symptom severity corresponding to slightly higher response rates ($\beta = 0.17$, $t(124) = 2.02$, $p = 0.046$). No other model steps accounted for significantly increased variance in either sample (S1: $R^2_{\text{change}} < .026$, $p > .26$; S2: $R^2_{\text{change}} < .078$, $p > .164$).

Overconfidence Results

In parallel with the contingency task, we conducted four hierarchical linear regressions predicting overestimation, overplacement, and both forms of overprecision. With the exception of one significant negative effect of the IDD-L on overplacement, current and lifetime depressive symptoms were not significantly associated with any overconfidence measure in Sample One nor in Sample Two (see Tables S8-S11). Across the overconfidence models, there was no consistent evidence of curvilinear effects, and inclusion of polynomial terms did not substantively change the nature of any overconfidence findings (see the Supplement for full results).

Discussion

In a well-powered, pre-registered study, we aimed to test the depressive realism hypothesis using the latest method-

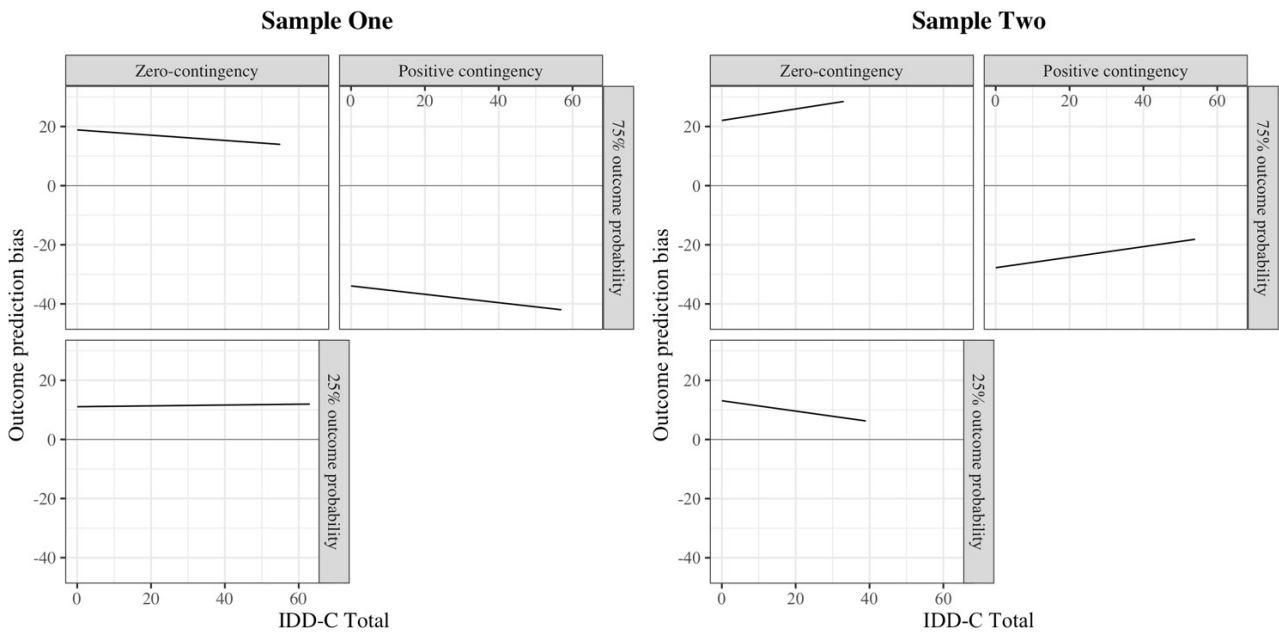


Figure 4. The Effect of IDD-C Score on Outcome Prediction Bias Across Contingency Conditions

ologies. We refined the contingency task to provide objective indices of control and to provide more measures of subjective control. Our study supplemented the original task with a well-validated measure of overconfidence, and included continuous indices of depressive severity, lifetime depression, hypomanic tendencies, anxiety, and trauma history.

Across two samples, we find no support for the depressive realism hypothesis. We first analyzed depressive realism using a standard, end-of-task measure eliciting control estimates. In our first sample, results contradicted depressive realism: participants with higher depressive symptoms displayed the greatest illusory control. This result appears to be driven by anxiety symptoms. In our second sample, depression symptoms did not significantly predict illusory control. In conditions with no contingency between response and outcome, participants tended to overestimate control in post-task ratings; higher anhedonic depression did not influence this illusory control. As expected, in circumstances with high actual control, they underestimated control (Gino et al., 2011); those with higher anhedonic depression displayed the same underestimation as did others. That is, estimates of control were noisy, regardless of current or lifetime depression. When we considered more frequent ratings of subjective control, we still found no association between depressive symptoms with outcome bias. Past work has suggested that psychomotor retardation in depression could be one mechanism through which depressive symptoms contribute to lower illusory control (Blanco et al., 2012), but we found no support for this hypothesis when examining frequency of responding in our samples.

Results of the overconfidence task mirror this null effect, in that current depressive symptom severity was not associated with *any* form of overconfidence. This not only contradicts the depressive realism hypothesis, but also con-

tradicts past work reporting decreased overestimation in depressed individuals (Fu et al., 2005, 2012; Stone et al., 2001). The possibility remains that other measures of depressive symptoms or other measures of perceived control could demonstrate such a relationship, but that would raise the question of how to conceptually reconcile differences with the two measures of depression and multiple measures of control that we used.

We explored a number of clinical features that could influence the effect of depression symptoms on illusory control and/or overconfidence. Our results suggest that anxiety may influence illusory control, supporting previous suggestions to consider anxiety symptoms when testing depressive realism (Ackermann & DeRubeis, 1991). We also examined the potential effect of levels of mental-health related impairment, current and lifetime diagnoses of or treatment for depression and anxiety, hypomanic tendencies, and trauma history on depressive realism. None of these variables showed a robust influence on either illusory control or overconfidence.

Our results must be considered in light of a number of potential methodological limitations. First, as with any behavioral task, it is possible that our null results may be best attributed to measurement error. We took two steps to address this concern. First, we included not only the original task, but also included a well-validated overconfidence task (D. A. Moore & Healy, 2008) to consider various domains in which depressive realism may manifest. Second, to assess measurement error as a potential cause for our null results, we tested mixed models controlling for trial-level random effects in the contingency task, a statistical approach which has been suggested to be one of the best available tools to account for measurement error (Parsons et al., 2019). Across these measurement and statistical approaches, and across two samples, we find no evidence that

depressive symptoms are tied to greater realism. Of course, it remains the case that we may not have fully recovered true effects (Rouder et al., 2019). Future research could further probe the reliability of the contingency task. It is possible that different tasks may be better suited to capturing the depressive realism effect.

While we acknowledge the innate limitations with sampling via MTurk, there is evidence that MTurk workers show more severe depression symptom severity than does the general population (Sheehan et al., 1996), and that their depression scores correspond to elevations in self-reported use of antidepressant medications (Ophir et al., 2019). We find remarkably consistent results across our MTurk and undergraduate samples, bolstering our confidence in the generalizability of our findings. That said, one limitation is that we exclusively sample US participants. We also did not measure participants' racial/ethnic identities, income, or socioeconomic status, limiting our ability to characterize our sample, and thus to understand how sample characteristics may influence the generalizability of our results. Extending our study to participants located outside the US could add valuable insight into generalizability, but we regard it as unlikely that a three-way interaction between culture, control, and depression proves a robust moderator of the results we report.

It is also worth noting that due to our ability to pre-screen comparatively fewer undergraduate than MTurk participants, Sample One included more participants with moderate-to-severe depression symptoms than Sample Two. However, the original depressive realism study (Alloy & Abramson, 1979) was conducted within a subclinical undergraduate sample, casting doubt on sample selection as the cause for our failure to find depressive realism. Several other studies utilizing contingency tasks detect a depressive realism effect in dysphoric populations (Ackermann & DeRubeis, 1991). Similarly, one may criticize the reliance on self-rated symptoms rather than clinical interviews in

assessing depression, but it is important to note that previous meta-analytic results find a stronger depressive realism effect for self-rated than interview-rated symptoms (M. T. Moore & Fresco, 2012).

The “sadder but wiser” hypothesis argues that depression enhances accuracy in judgment (Alloy & Abramson, 1979, 1988). Our results contradict these claims. We do not find that depressive symptoms correlate consistently with assessments of personal control. The implication is that errors in the assessment of personal control or overconfidence do not appear to be integral to depression.

Contributions

Contributed to conception and design: DAM, ASD, SLJ

Contributed to acquisition of data: ASD, KTG

Contributed to analysis and interpretation of data: ASD, KTG with supervision from DAM, SLJ

Drafted and/or revised the article: ASD, DAM, SLJ, KTG

Approved the submitted version for publication: ASD, DAM, SLJ, KTG

Competing Interests

An author on this manuscript (initials: DAM) serves as an associate editor at *Collabra: Psychology*.

Data Accessibility Statement

Our materials, including survey instruments, data, analysis code, are posted publicly on the project's Open Science Framework page [here](#).

Submitted: August 29, 2022 PDT, Accepted: September 01, 2022 PDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Ackermann, R., & DeRubeis, R. J. (1991). Is depressive realism real? *Clinical Psychology Review*, *11*(5), 565–584. [https://doi.org/10.1016/0272-7358\(91\)90004-e](https://doi.org/10.1016/0272-7358(91)90004-e)
- Allan, L. G., Siegel, S., & Hannah, S. (2007). The sad truth about depressive realism. *Quarterly Journal of Experimental Psychology*, *60*(3), 482–495. <https://doi.org/10.1080/17470210601002686>
- Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, *108*(4), 441–485. <https://doi.org/10.1037/0096-3445.108.4.441>
- Alloy, L. B., & Abramson, L. Y. (1988). Depressive realism: Four theoretical perspectives. In *Cognitive Processes in Depression* (pp. 223–265). Guilford Press. <https://search.proquest.com/psycinfo/docview/617501399/BFED0ED91745403FPQ/1>
- Alloy, L. B., Albright, J. S., Abramson, L. Y., & Dykman, B. M. (1990). Depressive realism and nondepressive optimistic illusions: The role of the self. In R. E. Ingram (Ed.), *Contemporary psychological approaches to depression: Theory, research, and treatment* (pp. 71–86). Plenum Press.
- Anderson, S. F., & Maxwell, S. E. (2017). Addressing the “Replication Crisis”: Using Original Studies to Design Replication Studies with Appropriate Statistical Power. *Multivariate Behavioral Research*, *52*(3), 305–324. <https://doi.org/10.1080/00273171.2017.1289361>
- Baker, A. G., Msetfi, R. M., Hanley, N., & Murphy, R. A. (2012). Depressive realism? Sadly not wiser. In M. Haselgrove & E. Hogarth (Eds.), *Clinical applications of learning theory* (pp. 153–177). Psychology Press.
- Blanco, F., Matute, H., & Vadillo, M. A. (2009). Depressive Realism: Wiser or Quieter? *The Psychological Record*, *59*(4), 551–562. <https://doi.org/10.1007/bf03395681>
- Blanco, F., Matute, H., & Vadillo, M. A. (2012). Mediating role of activity level in the depressive Realism effect. *PLOS ONE*, *7*(9), e46203. <https://doi.org/10.1371/journal.pone.0046203>
- Bucklin, S. (2017, June 22). Depressed people see the world more realistically. *Vice*. https://www.vice.com/en_us/article/8x9j3k/depressed-people-see-the-world-more-realistically
- Burkeman, O. (2006, November 25). Oliver Burkeman: Depressive realism. *The Guardian*. <https://www.theguardian.com/lifeandstyle/2006/nov/25/healthandwellbeing.features1>
- Burton, N. (2012, June 5). Depressive Realism: Wisdom or madness? *Psychology Today*. <http://www.psychologytoday.com/blog/hidden-and-see/201206/depressive-realism>
- Dobson, K. S., & Pusch, D. (1995). A test of the depressive realism hypothesis in clinically depressed subjects. *Cognitive Therapy and Research*, *19*(2), 179–194. <https://doi.org/10.1007/bf02229693>
- Dunning, D., & Story, A. L. (1991). Depression, realism, and the overconfidence effect: Are the sadder wiser when predicting future actions and events? *Journal of Personality and Social Psychology*, *61*(4), 521–532. <http://doi.org/10.1037/0022-3514.61.4.521>
- Eckblad, M., & Chapman, L. J. (1986). Development and validation of a scale for hypomanic personality. *Journal of Abnormal Psychology*, *95*(3), 214–222. <http://doi.org/10.1037/0021-843x.95.3.214>
- Everaert, J., Koster, E. H. W., & Derakshan, N. (2012). The combined cognitive bias hypothesis in depression. *Clinical Psychology Review*, *32*(5), 413–424. <https://doi.org/10.1016/j.cpr.2012.04.003>
- Feltham, C. (2017). Depressive realism: Interdisciplinary perspectives. In *Depressive realism: Interdisciplinary perspectives*. Routledge/Taylor & Francis Group.
- Fu, T., Koutstaal, W., Fu, C. H. Y., Poon, L., & Cleare, A. J. (2005). Depression, confidence, and decision: Evidence against depressive realism. *Journal of Psychopathology and Behavioral Assessment*, *27*(4), 243–252. <https://doi.org/10.1007/s10862-005-2404-x>
- Fu, T., Koutstaal, W., Poon, L., & Cleare, A. J. (2012). Confidence judgment in depression and dysphoria: The depressive realism vs. negativity hypotheses. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*(2), 699–704. <https://doi.org/10.1016/j.jbtep.2011.09.014>
- Gino, F., Sharek, Z., & Moore, D. A. (2011). Keeping the illusion of control under control: Ceilings, floors, and imperfect calibration. *Organizational Behavior and Human Decision Processes*, *114*(2), 104–114. <https://doi.org/10.1016/j.obhdp.2010.10.002>
- Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: Current status and future directions. *Annual Review of Clinical Psychology*, *6*(1), 285–312. <https://doi.org/10.1146/annurev.clinpsy.121208.131305>
- Konnikova, M. (2014, June 18). Don't Worry, Be Happy. *The New Yorker*. <https://www.newyorker.com/science/maria-konnikova/dont-worry-be-happy>
- Lewinsohn, P. M., Joiner, T. E., & Rohde, P. (2001). Evaluation of cognitive diathesis-stress models in predicting major depressive disorder in adolescents. *Journal of Abnormal Psychology*, *110*(2), 203–215. <http://doi.org/10.1037/0021-843x.110.2.203>
- Marchetti, I., Everaert, J., Dainer-Best, J., Loeys, T., Beevers, C. G., & Koster, E. H. W. (2018). Specificity and overlap of attention and memory biases in depression. *Journal of Affective Disorders*, *225*, 404–412. <https://doi.org/10.1016/j.jad.2017.08.037>
- Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annual Review of Clinical Psychology*, *1*(1), 167–195. <https://doi.org/10.1146/annurev.clinpsy.1.102803.143916>
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. <https://doi.org/10.1037/0033-295x.115.2.502>

- Moore, M. T., & Fresco, D. M. (2012). Depressive realism: A meta-analytic review. *Clinical Psychology Review*, 32(6), 496–509. <https://doi.org/10.1016/j.cpr.2012.05.004>
- Murphy, R. A., Vallée-Tourangeau, F., Msetfi, R., & Baker, A. G. (2005). Signal-outcome contingency, contiguity, and the depressive realism effect. In A. J. Wills (Ed.), *New directions in human associative learning* (pp. 193–219). Lawrence Erlbaum Associates Publishers.
- Ophir, Y., Sisso, I., Asterhan, C. S. C., Tikochinski, R., & Reichart, R. (2019). The Turker blues: Hidden factors behind increased depression rates among Amazon's Mechanical Turkers. *Clinical Psychological Science*, 8(1), 65–83. <https://doi.org/10.1177/2167702619865973>
- Parsons, S., Kruijt, A.-W., & Fox, E. (2019). Psychological Science Needs a Standard Practice of Reporting the Reliability of Cognitive-Behavioral Measurements. *Advances in Methods and Practices in Psychological Science*, 2(4), 378–395. <https://doi.org/10.1177/2515245919879695>
- Prisciandaro, J. J., & Roberts, J. E. (2005). A taxometric investigation of unipolar depression in the national comorbidity survey. *Journal of Abnormal Psychology*, 114(4), 718–728. <https://doi.org/10.1037/0021-843x.114.4.718>
- Raven, J. C. (1963). *Mental tests used in genetic studies: The performance of related individuals on tests mainly educative and mainly reproductive* [Unpublished master's thesis]. University of London.
- Rogers, W. H., Adler, D. A., Bungay, K. M., & Wilson, I. B. (2005). Depression screening instruments made good severity measures in a cross-sectional analysis. *Journal of Clinical Epidemiology*, 58(4), 370–377. <https://doi.org/10.1016/j.jclinepi.2004.10.010>
- Rouder, J., Kumar, A., & Haaf, J. M. (2019). Why Most Studies of Individual Differences With Inhibition Tasks Are Bound To Fail. *PsyArXiv*. <https://doi.org/10.31234/osf.io/3cjr5>
- Rude, S. S., Valdez, C. R., Odom, S., & Ebrahimi, A. (2003). Negative cognitive biases predict subsequent depression. *Cognitive Therapy and Research*, 27(4), 415–429. <https://doi.org/10.1023/a:1025472413805>
- Segal, Z. V., Kennedy, S., Gemar, M., Hood, K., Pedersen, R., & Buis, T. (2006). Cognitive reactivity to sad mood provocation and the prediction of depressive relapse. *Archives of General Psychiatry*, 63(7), 749–755. <https://doi.org/10.1001/archpsyc.63.7.749>
- Sheehan, D. V., Harnett-Sheehan, K., & Raj, B. A. (1996). The measurement of disability. *International Clinical Psychopharmacology*, 11(Supplement 3), 89–95. <https://doi.org/10.1097/00004850-199606003-00015>
- Soderstrom, N. C., Davalos, D. B., & Vázquez, S. M. (2011). Metacognition and depressive realism: Evidence for the level-of-depression account. *Cognitive Neuropsychiatry*, 16(5), 461–472. <https://doi.org/10.1080/13546805.2011.557921>
- Stern, G. S., & Berrenberg, J. L. (1979). Skill-set, success outcome, and mania as determinants of the illusion of control. *Journal of Research in Personality*, 13(2), 206–220. [https://doi.org/10.1016/0092-6566\(79\)90031-x](https://doi.org/10.1016/0092-6566(79)90031-x)
- Stone, E. R., Dodrill, C. L., & Johnson, N. (2001). Depressive cognition: A test of depressive Realism versus negativity using general knowledge questions. *The Journal of Psychology*, 135(6), 583–602. <https://doi.org/10.1080/00223980109603722>
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103(2), 193–210. <https://doi.org/10.1037/0033-2909.103.2.193>
- Taylor, S. E., Lerner, J. S., Sage, R. M., Lehman, B. J., & Seeman, T. E. (2004). Early environment, emotions, responses to stress, and health. *Journal of Personality*, 72(6), 1365–1394. <https://doi.org/10.1111/j.1467-6494.2004.00300.x>
- Venkatesh, S. (2013). *An Investigation of the Depressive Realism Effect and Its Mechanisms*. <http://unsworks.unsw.edu.au/fapi/datastream/unsworks:11776/SOURCE01?view=true>
- Venkatesh, Shruti, Moulds, M. L., & Mitchell, C. J. (2018). Testing for Depressive Realism in a Clinically Depressed Sample. *Behaviour Change*, 35(2), 108–122. <https://doi.org/10.1017/bec.2018.12>
- Wardenaar, K. J., van Veen, T., Giltay, E. J., de Beurs, E., Penninx, B. W., & Zitman, F. G. (2010). Development and validation of a 30-item short adaptation of the Mood and Anxiety Symptoms Questionnaire (MASQ). *Psychiatry Research*, 179(1), 101–106. <https://doi.org/10.1016/j.psychres.2009.03.005>
- Wong, K. F. E., Yik, M., & Kwong, J. Y. (2006). Understanding the emotional aspects of escalation of commitment: The role of negative affect. *Journal of Applied Psychology*, 91(2), 282–297. <https://doi.org/10.1037/0021-9010.91.2.282>
- Zetsche, U., Bürkner, P.-C., & Renneberg, B. (2019). Future expectations in clinical depression: Biased or realistic? *Journal of Abnormal Psychology*, 128(7), 678–688. <https://doi.org/10.1037/abn0000452>
- Zimmerman, M., & Coryell, W. (1987). The inventory to diagnose depression, lifetime version. *Acta Psychiatrica Scandinavica*, 75(5), 495–499. <https://doi.org/10.1111/j.1600-0447.1987.tb02824.x>

Supplementary Materials

Peer Review History

Download: https://collabra.scholasticahq.com/article/38529-sadder-wiser-depressive-realism-is-not-robust-to-replication/attachment/101017.docx?auth_token=BwMSjdvUfmjgnaFoEUJg

Supplemental Material

Download: https://collabra.scholasticahq.com/article/38529-sadder-wiser-depressive-realism-is-not-robust-to-replication/attachment/101018.docx?auth_token=BwMSjdvUfmjgnaFoEUJg
