2022

Challenging Depressive Beliefs: Habitual and Recollective Components of Stability or Change

Paula T. Hertel
Trinity University, phertel@trinity.edu

M. C. Acuff

J. Hernandez

E. Poppe

Publication Details
Journal of Behavior Therapy and Experimental Psychiatry

Challenging Depressive Beliefs:

Habitual and Recollective Components of Stability or Change

Paula Hertel

M. Chase Acuff

Jessica Hernandez

Elizabeth Poppe

Maybe not the final version

In press in 2022: *Journal of Behavior Therapy and Experimental Psychiatry*.

Abstract

Background and objectives. Depressed people tend to hold stable negative beliefs that resist challenges. Two experiments investigated the cognitive bases of belief change or resistance to change following the provision of supportive or challenging pseudo-evidence.

Method. Students scoring high and low on a measure of depressed state read belief statements, each followed by invented experimental evidence to either verify or discount them. Two days later, they read all the belief statements again, together with new statements, this time rating belief.

Results. The students agreed that the statements described common beliefs and that the evidence was plausible. Discounted statements were believed less than new statements on the test. Also, dysphoric students believed discounted and new statements less than verified statements, but that difference was larger for the nondysphoric students. Parameter estimates of the habitual basis for belief ratings, obtained with process-dissociation procedures, were higher in the dysphoric group, and estimates of evidence recollection were lower. The latter finding was conceptually supported by deficient recognition of the gist of the discounting evidence in the dysphoric group (Experiment 2).

Limitations. Experiment 2 results replicated the rating effects in Experiment 1, but not the parameter differences, due to low power as a consequence of the university response to the pandemic.

Conclusions. We interpret these results in the context of other evidence regarding belief change and depressive cognition, such as habitual rumination and deficient cognitive control.

Key Words: belief updating, misinformation correction, depression, habit, recollection

Challenging Depressive Beliefs:

Habitual and Recollective Components of Stability or Change

The research we report focuses on common beliefs of an emotionally negative nature—things that people, particularly depressed people, might say to themselves to sum up their situation or the state of the world; for example: *People are generally unreliable*. These are the kinds of beliefs that psychotherapists hope will change as they help their clients become less pessimistic and stultified. But as Abramson et al. (1989) cautioned long ago, these stable and pervasive beliefs are not so easily changed. One possible reason is that people experiencing depression tend habitually to think about their distorted perceptions as they ruminate (Watkins & Nolen-Hoeksema, 2014), and the mere repetition of assertions, true or false, increases their truth value (Hasher et al., 1977). In clinical theory, these habitual thoughts reflect depressive schemas that both reflect and influence perceptions of self, the world, and the future (Beck, 1967; see LeMoult & Gotlib, 2019). The purpose of our experiments was not to add to our understanding of the etiology of negative beliefs in depression, however. Instead, inspired by a nonclinical cognitive literature, we used the lens of a processing framework to focus on mechanisms underlying the success or failure of attempts to modify those beliefs.

Challenging beliefs is the topic of a research tradition in which investigators try to change faulty beliefs through the correction of misinformation (such as fake news; see Lewandowsky et al., 2017). The continued influence effect (Johnson & Seifert, 1994), emerging from experiments containing such challenges, refers to the fact that informative challenges rarely vanquish false beliefs entirely. After a variety of "interventions" for example, some people still tend to belief that vaccinations harm children, regardless of evidence to the contrary. Nevertheless, at least short-term success in diminishing faulty beliefs has been achieved (e.g., Wahlheim et al., 2020; see the reviews by Lewandowsky et al., 2012, and Rapp, 2016), with individual differences in the degree of such success comprising the error term.

Although the type of ideas put up for change are different in the two research domains—negative beliefs on the one hand and political "facts" and conspiracies on the other—both types are difficult to change. In the case of negative beliefs, however, speculation about the nature of individual differences is straightforward. People who show signs of depression should show greater resistance to change, whether we express this resistance in terms of stable mental schemas or inflexible cognitive processes (e.g., Everaert et al., 2018; Stange et al., 2017). Depression is also characterized by habitual rumination and repetitive thinking (Watkins & Nolen-Hoeksema, 2014) and deficiencies in cognitive control (Hertel, 2004), control being crucial for recalling disconfirming evidence. The experiments in this report were designed to investigate these latter two possible bases for resistance to belief change in young and nonclinical samples.

As we considered procedures for causing change, it occurred to us that people are rarely exposed to the sort of empirical evidence concerning depressive beliefs typically used to challenge other types of faulty beliefs. Such evidence would be hard to find, given the global and abstract nature of many depressive beliefs. We therefore invented it. After collecting beliefs that seem to fit a depressive mindset, we concocted evidence to verify or discount each belief. Our goals were: (a) to discover whether students who self-report depressive symptoms would show greater resistance to the effects of discounting evidence and (b) to investigate the habitual and recollective bases for depression-related differences in belief change. Next, we describe our approach.

Bases of Belief Change

Because no task is process pure (Jacoby et al., 1989), investigators can reason that making decisions about belief involves, in general terms, two types of processes—one type that represents our habitual tendencies to respond in a certain way and another type that is more analytic in terms of the deliberate search for evidence that might weigh on the decision. If the task is to decide about whether

you have studied something earlier, for example, both the automatic or habitual aspects of recruiting the prior experience and the recollective processes that summon contextual evidence join together to influence the response (Jacoby, 1991). Judgments about belief can be thought about similarly (see Begg et al., 1992). By using Jacoby's general scheme for process dissociation, we can imagine and even construct situations in which successful habitual and recollective processes operate together and others in which they operate in opposition. To wit, on some trials in our experiment the habitual tendency to believe a depressive statement (H for habit) and to remember that empirical research has verified the statement (R for recollection) jointly lead to a positive response of endorsing the statement as true. In other words, the probability of believing a statement that has been verified is a function of the probability of recollecting the verifying evidence (*R*) or, in the absence of recollection (1-*R*), the probability that one habitually believes such statements (*H*); stated mathematically:

$$P\,(belief|verified) = R + (1 - R)H \qquad \text{[Equation 1]}$$

On the other hand, if the statement has been discounted, the two types of processes work in opposition, in the sense that if the discounting evidence is recollected, the statement would not be believed. Therefore, if the discounted statement *is* believed, that endorsement would be based on habit (*H*) in the absence of evidence recollection (1-*R*); mathematically:

$$P\,(belief|discounted) = \quad (1 - R)H \qquad \text{[Equation 2]}$$

These equations are useful means of stating the assumptions, because if Equation 2 is subtracted from Equation 1 (after substituting the proportion of statements believed for the probability of belief), then we have a way to estimate recollective processes involved in belief change:

$$R = \text{Proportion believed}_{\text{verified trials}} - \text{Proportion believed}_{\text{discounted trials}} \qquad \text{[Equation 3]}$$

Subsequently, habit can be estimated by substituting the estimate of recollection into Equation 2.

$$H = \text{Proportion believed}_{\text{discounted trials}} / (1 - R) \qquad \text{[Equation 4]}$$

How should these estimates differ according to the individuals' depressive status? Clearly, habit should be stronger for depressed people, given their presumed tendencies to hold such beliefs about themselves, the world, and the future, and to entertain them occasionally. *H*, the estimate of habitual negative belief, should be higher in our dysphoric group. (See evidence concerning resistance to disconfirmation in the reviews by Kube & Rozenkrantz, 2021; Stange et al., 2017.) Moreover, evidence also indicates that performance on tasks that rely on cognitive control is typically impaired in depressed states (see the review by LeMoult & Gotlib, 2019). Therefore, estimates of recollection (*R*) might be lower for our dysphoric participants, either because they would initiate fewer attempts to retrieve the evidence as they rate belief or because they actually recollect it less well. One reason to waffle on the latter prediction about deficient estimates of *R* is that our participants were students selected according to self-reported depressive symptoms. Evidence for cognitive deficits associated with dysphoria (non-diagnosed depressive states) is not as reliable (Gotlib & Joormann, 2010).

Summary and Predictions

This project focused on the bases of belief change that might reveal depression-related differences. Borrowing from the experimental tradition of correcting misinformation, we discounted some negative belief statements (while verifying others) and examined the consequences on ratings of beliefs in those statements made two days subsequently. We also included new belief statements as controls in the rating task. (Thus, statement condition served as a within-subjects factor, comprising verified, discounted, and control statements.)

Our first set of predictions addressed belief ratings. we predicted that beliefs would be affected by the evidence; this effect would be revealed by lower ratings for discounted statements, compared to

control statements. We also predicted that the dysphoric students would believe the discounted statements more than would the other students. And the dysphoric group should also produce elevated ratings for the control statements, presented for the first time at test. This latter difference suggests that the discounting effect (the difference in ratings between control and discounted belief statements) might not differ according to the participants' category. Therefore, the interaction was not predicted, and neither was a group difference in ratings for verified statements. Verified statements were included in the design in order to potentiate process-dissociation procedures and to establish variety in the evidence, so that a general strategy to accept or doubt evidence would not emerge.

By using process-dissociation procedures, Experiment 1 focused on the mechanisms of belief change, instantiated as estimates of habitual belief or belief based on recollecting the evidence. We predicted that group differences in post-evidential beliefs would be understood as differences in habitual negative beliefs (*H*) and possibly also as differences in the tendency or ability to recollect the discounting evidence (*R*). Habitual belief should remain stronger for dysphoric participants (as it would be in the absence of evidence), and the recollective basis of their belief judgments might be weaker.

Experiment 1

Method

Participants

Our stopping rule for data collection was determined by our two-semester pool of qualifying students who were willing to participate. When the limit was neared, we recruited the largest cell size possible while preserving the complete counterbalancing of statements with evidence condition (see the section on Materials). First, we screened all students attending introductory-psychology classes at our university by administering the Beck Depression Inventory (BDI-II, missing the suicide item; Beck, Steer, & Brown, 1996). Within a few days or as long as three weeks after screening, we recruited

students whose BDI-II scores fell in the first and fourth quartiles; 58 agreed to participate and came to the lab, without any connection to screening having been mentioned. At the end of Session 1 a second BDI-II was administered, and we did not examine the belief data produced by students whose scores on that administration deviated from the initial quartile cutoffs (6 and 17) in the "wrong" direction by more than 2 points.[1] While observing the constraint of equal cell size, we randomly assigned participants to a counterbalancing condition. When students' second BDI-II score disqualified their data (four initially dysphoric students), they were each replaced by another student in the quartile pool, so that complete counterbalancing was achieved. Six students (five dysphoric) did not participate in Session 2, and they were replaced similarly. The final sample consistent of 24 dysphoric and 24 nondysphoric students. The average age was 18.9 years. The students identified primarily as White/Caucasian (67%), Hispanic/Latinx (15%), East or Southest Asian (9%), Black/African-American (5%), Pacific Islander (4%), and one student identified as Native American.

Materials

In a previous semester, we recruited students enrolled in an introductory psychology course who had previously scored in the first and fourth quartiles of the BDI-II. They each read a subset of 100 statements and judged the extent of their agreement with each statement on a 9-pt scale, ranging from "strongly disagree" (1) to "strongly agree" (9). Statements were obtained from a variety of published sources and internet posts and especially from a set of belief statements used by Van Boekel et al. (2017). Most statements represented abstract negative beliefs; others were the positive version of some of the negative beliefs, and still others were neutral. The survey ended with another administration of the BDI-II. For Experiment 1 in this report, we chose 18 negative statements for which the pilot ratings

[1] In one exception, an initially dysphoric student participated after a long delay and scored low on the BDI-II, and so we included her data in the nondysphoric group.

were higher for the dysphoric, compared to the non-dysphoric participants in this pilot sample (e.g., "The world is an unforgiving place"). Then we chose 6 positive statements (e.g., "People often look out for each other") and 6 neutral fillers that in our view added variety (e.g., "Lie detectors do not provide quality evidence for criminal prosecution") to discourage response strategies. Materials are available at https://osf.io/rs4za/.

We distributed the negative statements into three sets of six, in a manner that balanced the sets on the mean and standard deviation of their pilot ratings and on the median difference between the ratings from the first- and fourth-quartile participants in the pilot sample. The three sets were rotated across the three experimental conditions—verified, discounted, and control (new statements at test)—to achieve complete counterbalancing of the statements with their experimental role. We constructed both versions of each research summary so that they would primarily differ in terms of verification or disconfirmation alone. To evaluate the plausibility of the evidence (and other aspects of method), we posed questions to participants at the end of the lab session in Experiment 1 (see Table 1).

The final program included six positive and six neutral statements; three of the positive and three of the neutral statements were assigned to be verified. The other three neutral statements were discounted, and the other three positives were reserved to be new controls on the text. Only the sets of negative statements were counterbalanced with their experimental role.

Procedure

Session 1. First, we told the participants about what they would encounter: Each trial contained a statement that describes a common belief, followed by a research summary describing empirical evidence about that belief. The experimenters' script was designed to maximize the credibility of our concocted research summaries (see Van Boekel et al., 2017). Experimenters carefully followed this

script (reproduced in the appendix), having practiced it thoroughly to establish a convincing demeanor and reduce skepticism.

Belief statements, each followed by a research summary, appeared in three blocks of seven and were randomized anew within blocks for each participant. Each block contained two statements from the set assigned for verification, two assigned for disconfirmation, one positive, and two filler. The statements were each presented for 6 s, and offset with the appearance of the research summary. At their own pace, having been asked to think carefully about the summary, participants pressed the spacebar to reveal a new screen asking whether the research summary verified or disconfirmed the prior statement. They pressed the "1" key to indicate the belief had been verified and "2" to indicate that it had been discounted. (Belief ratings were analyzed contingent upon the correctness of this decision.) The experimenter led the participant through two examples with additional filler materials before starting the main program and moving aside.

At the end of Session 1, we asked the participants to fill out a survey to provide feedback about the experiment. We asked them to be very honest and told them that their feedback would be very helpful to us in determining "things we might need to change for next semester." The items are listed in Table 1, and the response scale ranged from 1 ("not at all") to 7 ("very much so"). Then we administered the self-paced BDI-II. (Anonymity was stressed throughout all phases of this project.) Mean BDI-II scores for both experiments are reported in Table 1. At the end of the session, the experimenter reminded participants to check email in two days for a link to additional statements to evaluate in the second part of the experiment. We told them that the break was necessary because the belief statements tend to blur together in our minds if we consider too many at the same time.

Session 2. Two days after the lab session, participants received an email with the link to the program that assessed their beliefs. The order loosely replicated block membership from Session 1, with

control statements added and fillers omitted. One statement from each of the three sets (one verified, one discounted, and one control) and one positive statement were assigned to each of six blocks, with order randomized anew for each participant. Three of the positive statements (shown in alternating blocks) had been previously presented and verified, and three were new controls on the test.

Instructions for rating belief stated that some of the belief statements had been shown in Session 1 and some were new; regardless, participants should rate how much they *currently* believed each one. Belief was rated on an 8-pt scale, anchored by "strongly disbelieve" and "strongly believe." A short rest occurred halfway. After the rating task, the BDI-II was again administered. Participation credit was contingent on completion of the online session within six hours of receiving the link and a return to the lab at some point during the same week to be debriefed.

Session 3. Within a few days of Session 2, participants returned to the lab for debriefing and credit. The debriefing emphasized the completely fake nature of the summaries of evidence and inquired about prior knowledge of that fact. (No one revealed it.) The experimenters provided instructions about consulting the university counseling service if needed.

Results and Discussion

Accuracy of Session-1 Responses

In Session 1, participants decided if each research summary verified or discounted the preceding belief statement. For each participant, the percentages of negative statements that were judged accurately were submitted to an analysis of variance with a between-subjects factor for group (dysphoric or other) and a within-subjects factor for summary type (verifying and discounting). Both summary-related effects in the overall design were nonsignificant, $p > .50$, $\eta_p^2 < .02$. The main effect of group was also nonsignificant and small, $p = .25$, $\eta_p^2 = .03$. Therefore, as illustrated by the means reported in the

top section of Table 2, the initial processing differences were nonsystematic with respect to group and the type of evidence.

Belief Ratings during Session 2

For each participant, we calculated the mean ratings for verified and discounted negative belief statements, contingent on accurate responding in Session 1, and we calculated the mean rating for the control statements that were new on the test. These means were submitted to a mixed-design analysis of variance with factors for group and statement condition (verified, discounted, and control). Cell means and standard deviations are reported in Table 2. Both main effects were significant and large. As expected, the dysphoric students believed the negative statements to a greater extent than did the other students, regardless of the type of evidence or their prior occurrence in Session 1, $F(1, 46) = 17.74$, $p < .001$, $\eta_p^2 = .29$. And regardless of group, the students' ratings depended on the condition assigned to the statements, $F(2, 92) = 42.97$, $p < .001$, $\eta_p^2 = .48$. The variance for this main effect was partitioned into a comparison to evaluate the significance and size of a discounting effect, plus the remaining orthogonal comparison. First, discounted statements were believed less strongly than control statements, $F(1, 46) = 11.47$, $p = .001$, $\eta_p^2 = .20$. Second, verified statements were believed more strongly than discounted and control statements, combined, $F(1, 46) = 88.43$, $p < .001$, $\eta_p^2 = .66$.

Although the overall interaction of group with statement condition was not significant, $F(2, 92) = 2.02$, $p = .14$, $\eta_p^2 = .04$, we evaluated the interaction of group with each of the comparisons described above. As we had anticipated, the rating difference between discounted and control statements did not depend on group, $p > .44$, $\eta_p^2 = .01$. However, the dysphoric students' ratings showed a smaller difference between verified statements and the other two conditions ($M_{dysphoric} = 1.1$ vs. $M_{others} = 1.7$), $F(1, 46) = 4.10$, $p = .049$, $\eta_p^2 = .08$.

Although few statements were classified as positive (only 3 verified and 3 new controls, we submitted ratings for new positive and verified positive statements to an analysis of variance with a between-subjects factor for group. This analysis revealed a significant main effect of group, $F(1, 46) = 42.32$, $p < .001$, $\eta_p^2 = .48$, and a nonsignificant interaction, $p = .183$, $\eta_p^2 = .04$. The mean belief ratings for new positive statements was 4.6 by dysphoric students and 5.7 by the others; the corresponding means for verified positive statements were 5.8 and 6.5. This pattern shows that dysphoric students' ratings were not higher for all statements and thereby subject to the interpretation of response bias.

Estimates of Habit and Recollection as Bases for Belief Ratings

Ratings for the verified and discounted statements were separately converted to proportion believed, contingent on correct responding to the research summary in Session 1 (the number of statements rated 5-8, divided by the number with previously understood evidence). Means are reported at the bottom of Table 2 and show the same patterns of significant effects as the raw ratings. As planned, the proportions of statements believed were converted to estimates of recollection and habit.

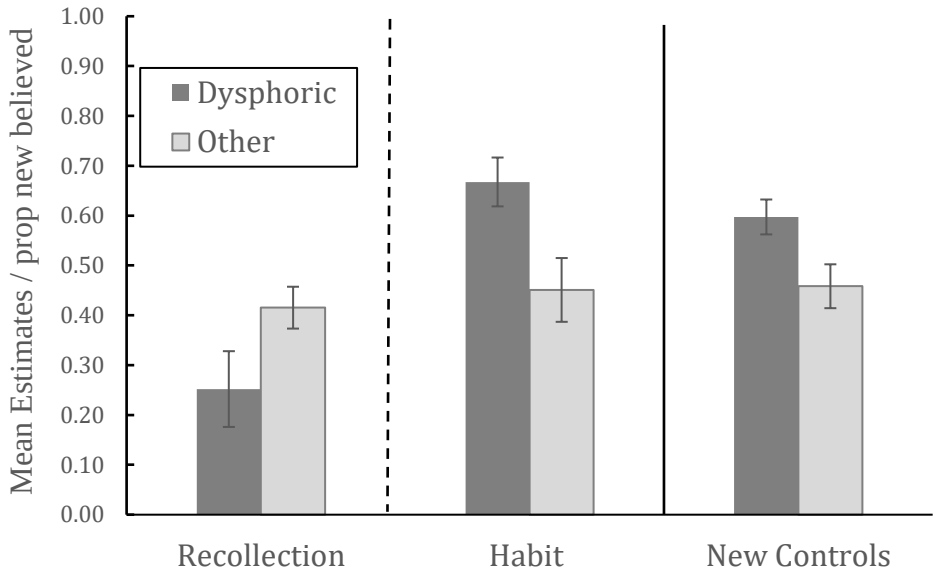
In the first step, we calculated estimates of recollection (R) for each individual by subtracting the proportion of believed discounted statements from the proportion believed verified statements (Equation 3). Estimates of habit (H) were obtained by substitution of R into Equation 2 (see Equation 4). The primary hypothesis about the bases for belief ratings was supported: Habit estimates (H) were significantly larger for the dysphoric students ($M = .67$ vs. $M = .45$ for others), $t(46) = 2.68$, $SE_{DIFF} = .081$, $p = .01$, Cohen's d = .78. The similarity of the habit estimates to the proportion of new control statements believed ($M = .60$ by dysphoric students vs. $M = .46$ by the others) informally provides converging evidence for using process dissociation equations to estimate habit.

Next, we examined the possible group difference in controlled recollection. Dysphoric students produced lower values of R ($M = .25$ vs. $M = .42$), as revealed by a nonsignificant trend in a two-tailed

test, $t(35.65) = 1.89$, $SE_{DIFF} = .087$, $p = .067$, Cohen's d = .57. (The justifiable one-tailed test was clearly significant.) Figure 1 depicts these differences.

Figure 1

Mean Estimates of Habit and Recollection and Mean Proportion of New Negative Statements Believed (Standard Error Bars)



Experiment 2

Two main issues led to the second experiment. First, we became concerned that the second process-dissociation assumption (Equation 2) might not represent the cognition of the dysphoric students. The equation assumes that their endorsement of the statement (a rating above 4) would occur only in the absence of recollecting the discounting research summary. After conducting Experiment 1, we realized that participants might have remembered the research summaries and simply ignored them while rating belief, as if they were saying, "I know what the evidence says, but I believe the statement anyway." This stance might well characterize dysphoric students more than others. We did interrogate the debriefing ratings to notice that the dysphoric participants found the summaries slightly less

plausible (see Hinze et al., 2014), compared to the other group of students, and the dysphoric students reported being slightly less willing to change their beliefs based on the evidence, although neither of these group differences approached statistical significance. Regardless, self reports do not constitute the best basis for confidence in the estimates. And because we did not ask a separate question about the plausibility of *discounting* evidence, participants would not have been able to express their specific doubts. Thus, our main goal for Experiment 2 was to find alternative evidence about memory for the gist of the research summaries—to ask if belief, like many other cognitive phenomena in depression, is subject to deficits in controlled remembering and not merely a reflection of unwillingness to change. This issue is important not only for the interpretation of the recollection deficit in Experiment 1, but also for the interpretation of the difference in habit estimates, the computation of which substantially relies on the proper interpretation of Equation 2.

Our second goal was to replicate the Experiment-1 results concerning belief change. One way to calculate the needed sample size was to use the effect size for the group difference in the estimate of recollection, which suggested a cell size of 44 (with power set at .80, one tailed; G Power 3.0.10; Faul et al., 2007). As detailed below, we achieved only 24 usable sets of belief data in each group before the project was interrupted by the university's change to remote classes in Spring, 2020, due to the pandemic. Still, in the interest of disclosure, we report the full results in the online supplement. In summary of those results, we found the same patterns of means for all measures, including estimates of recollection and habit, but the estimate differences did not approach statistical significance and the effect sizes were smaller. In addition, not much confidence can be placed in these partial results, because counterbalancing was disturbed and the plausibility ratings were lower.

Method

Participants and Design

Our new design for Experiment 2 contained a between-subjects factor for when the recognition test was administered: Alone or after the test of belief change. Following the recognition test with the beliefs test seemed pointless, given the degree of reminders the recognition test would provide. Yet we also could not be sure that the beliefs test would not affect recognition performance, so we included the Recognition-Only group with a plan to test for the possible effect of the preceding belief test. This design was preregistered at https://osf.io/rs4za/, where all statements, summaries, and recognition items can be found.

From a pool similar to that in Experiment 1, we recruited 117 students to participate. Within the first and fourth quartiles that resulted from screening on the BDI-II, the students were randomly assigned to one of six between-subjects conditions, resulting from the crossing of testing condition with the counterbalancing factor for rotating statement sets across the statement conditions (verified, discounted, and new on the test).[2] The data from 27 students were set aside and replaced by recruiting other students from the corresponding quartile. From the 117, 11 original participants did not take the test or took it at least a day late, and 12 participants (8 dysphoric and 4 others) were replaced when their Session-1 BDI-II scores fell out of the original quartile by more than 2 points. The data from four additional students (the last ones participating in each relevant cell), were removed in order to make counterbalancing complete for the recognition test. The final samples included 24 in the Beliefs-First condition and 21 in the Recognition-Only condition, within each group (dysphoric and other), 90 students in total. The average age was 18.5 years. The students identified primarily as White/Caucasian (60%), Hispanic/Latinx (20%), East or Southeast Asian (10%), Black/African-American (4%), Middle Eastern (2%), and three students chose "other" or did not respond.

[2] This rotation also applied in the Recognition-Only condition, even though new items could not be included in the forced-choice recognition analysis.

Materials and Procedure

The materials and procedures were identical to those used in Experiment 1, with the exception of the forced-choice recognition test. We created two versions, reversing the positions on the screen for the two alternatives. (This position factor was loosely counterbalanced with the other between-subjects factors.) Each item largely reproduced the original summary but omitted any information that would bias the choice of the alternatives. For example:

> **Statement:** People are generally unreliable.
>
> **Verifying Summary:** When asked to compile information into a spreadsheet whenever they had time over the course of one week, only 19% of 349 participants entered the information accurately and before the deadline (Riley & Jacobs, 2006).
>
> **Discounting Summary:** When asked to compile information into a spreadsheet whenever they had time over the course of one week, 92% of 349 participants entered the information accurately and before the deadline (Riley & Jacobs, 2006).
>
> **Test Item**: When asked to compile information into a spreadsheet whenever they had time over the course of one week, _____ participants entered the information accurately and before the deadline (Riley & Jacobs, 2006).
>
> A. Few
> B. Many

The program instructed participants to choose the alternative that represents the meaning of the summary they read two days earlier. Items on the test followed the same randomized block order that was used for the test of beliefs in Experiment 1. The recognition test was self-paced, and it was followed by another administration of the BDI-II.

Results

Accuracy of Session-1 Responses

In Session 1, participants decided if each research summary verified or discounted the preceding belief statement. For each participant, the percentages of the six responses to negative statements that were judged accurately were submitted to an analysis of variance with between-subjects factors for

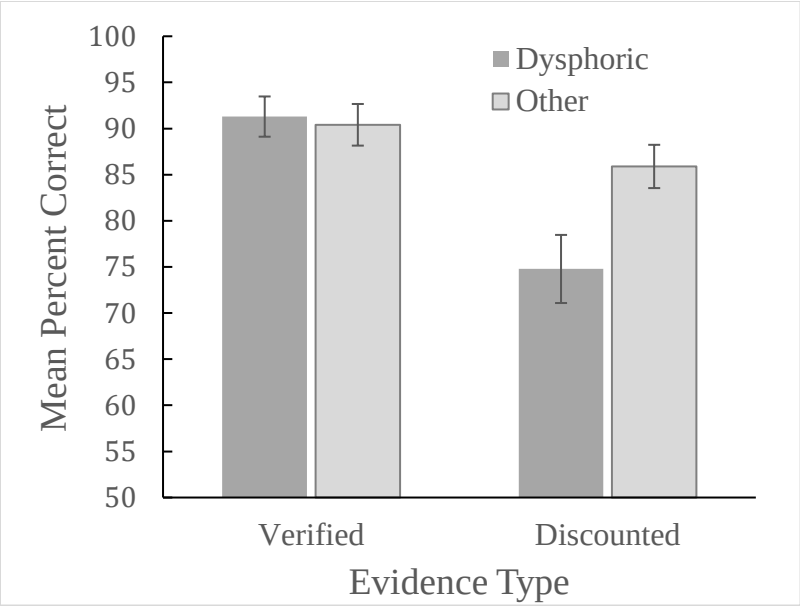
group (dysphoric or other) and testing condition (beliefs tested first or recognition only), and a within-subjects factor for evidence type (verifying and discounting). Most effects in the overall design were nonsignificant and small, $p > .29$, $\eta_p^2 < .02$; however, the main effect of group showed a nonsignificant trend toward slightly lower accuracy by dysphoric students ($M_D = 88.1$, $SD = 10.43$, $M_O = 91.5$, $SD = 8.24$), $F(1, 86) = 3.16$ $p = .08$, $\eta_p^2 = .04$.The means collapsed across testing condition are reported in the online supplement.

Forced Choice Recognition

The primary reason for conducting Experiment 2 was to examine recognition memory for the gist of the research evidence. The percentages of correct choices, conditional on accurate judgments in Session 1, were submitted to an analysis of variance with between-subjects factors for group (dysphoric or other) and testing condition (beliefs first or recognition only), and a within-subjects factor for evidence type (verifying and discounting). No effect associated with testing condition reached significance, $p > .16$, $\eta_p^2 < .02$. The significant main effect of evidence type, $F(1, 86) = 18.03$, $p < .001$, $\eta_p^2 = .17$, was qualified by the interaction of evidence type with group, $F(1, 86) = 6.26$, $p = .014$, $\eta_p^2 = .07$. Figure 2 depicts the means. Dysphoric students made more errors in recognizing the gist of discounting evidence, $t(74.54) = 2.54$, $SE_{DIFF} = 4.37$, $p = .013$, Cohen's d = .54. However, the mean recognition scores for verifying evidence did not significantly vary with group status, $p = .778$, Cohen's d = .06.

Figure 2

Mean Percentage Correct on the Forced-Choice Recognition Test (Standard Error Bars, n = 45.) Chance responding = 50%.



General Discussion

In our attempt to experimentally challenge depressive beliefs with evidence, we showed effects established by summaries of fake research read two days earlier. Summaries that discounted the negative belief statements significantly lowered belief ratings, in comparison to ratings for new belief statements, seen for the first time on the test. However, this discounting effect did not emerge as a deficit in the dysphoric group, because their ratings for the new statements were also higher. More generally, the evidence that dysphoria is associated with holding very general negative beliefs was strong, and the overall effects of verification and disconfirmation were large. Yet specific dysphoria-related differences could nevertheless be detected by evaluating two bases for the belief judgments: the continuation of habitual belief and recollection of evidence from the research summaries, bases placed in opposition to each other when the research summary discounted the belief statement.

The large difference in the general tendency to endorse negative beliefs, regardless of evidence, itself was consistent with cognitive habits in depression. Similarly, the habit component of the rated beliefs can be understood procedurally as the result of their repeated conceptual replication in the mental experience of depressed people. However, habit is not the complete story behind the stability of negative beliefs shown by our dysphoric participants. In using process-dissociation procedures, we assumed that the endorsement of each belief statement would reflect the independent contributions of habitual belief in the statement and recollection of research evidence that either supported or challenged it. Estimates of habit were significantly and substantially higher in the dysphoric group. Estimates of recollection were significantly lower, as revealed by a one-tailed test—a test justified by a large literature on depressive related deficits in cognitive control (see LeMoult & Gotlib, 2019), some of which has been found by using other process-dissociation paradigms. Nevertheless, the possibility that Equation 2 did not well represent processes involved in belief judgments in particular led us to seek independent evidence about memory deficits. The dysphoric students may have remembered the discounting evidence as well as others did but simply ignored it in their fidelity to the negative beliefs. Therefore, in Experiment 2 we assessed recognition of the gist of the evidence, and we found that the dysphoric students recognized the gist of discounting evidence less well. This result is consistent with the assumption that lower recollection allowed habit to guide their belief judgments (see Equation 2).

As a separate issue, it is notable that the recognition deficit in Experiment 2 was not found for verifying summaries—summaries with quite similar content and structure as the discounting summaries. Therefore, the gist-recognition results do not fit neatly into the body of research findings concerning depressive deficits in cognitive control. Although a clear explanation cannot be determined from the design of Experiment 2, possibilities include reduced elaboration of discounting summaries during initial exposure and consequentially deficient recognition. This account is consistent with research on

deficient cognitive initiative (see Hertel & Rude, 1991) and motivation to not think about information that challenges depressive beliefs (see Kube & Rozenkrantz, 2021). Another possibility is that the dysphoric students suppressed the discounting evidence, hampering its further consideration. Similar valence-related imbalances in belief change have been observed in a variety of contexts (e.g., Sharot & Garrett, 2016). Our recognition results suggest that the contribution of memory phenomena to imbalanced changes in belief merits continued investigation.

Qualification and Limitations

One limitation in our work concerns the lack of power in Experiment 2 to adequately assess the replicability of differences in the parameter estimates. The similar pattern in the ratings and estimates across the two experiments, however, provides some basis for concluding that memory for the discounting evidence operated similarly across the experiments in affecting (or failing to affect) belief ratings. Regardless, future replication of the parameter differences would increase our confidence.

Any attempt at replication must consider our sense that successful demonstration of belief change in this paradigm is highly dependent on the acting ability of the experimenters; the delivery of the background and the instructions must be casual, yet truly convincing. As Experiment 2 proceeded, this was a matter for concern that possibly affected the belief ratings, and it convinced us that a continuation of that experiment with an online version of the task was not worth the effort. In support of that analysis, the debriefing ratings in Experiment 2 suggested that the dysphoric students found the evidence a little less plausible than did the others (see Table 1).

We also suggest replication with an older sample and by investigators with access to clinical samples. Stability in habits of belief should increase with age. And a clinical sample might express stronger beliefs, because students are taught to be skeptical of platitudes; in fact, our dysphoric students believed only slightly more than half of the control statements. On the other hand, the student samples

were informative, especially considering that, even though college students are well practiced in memory tasks, their recollective use of the evidence while judging beliefs was fairly low in both groups. This outcome suggests that they did not treat the belief-rating task solely as an attempt to measure memory, as they might have done, because gist recognition was surprisingly high when it was requested of the similar sample in Experiment 2.

The second alert we raise concerns conclusions about the role played by the new control statements. The control provided by the absence of challenges to the "misinformation" (negative beliefs, in our case) was confounded by the absence of prior exposure to the statements themselves (at least within our sessions). We chose not to present these "control" statements in Session 1 because we were worried about general carryover effects that could result from too many statements similar in tone. As a consequence, rating differences between the old and new statements on the test reflect differences in prior exposure as well as effects from the concocted evidence. Repetition increases belief (Hasher et al., 1977), and therefore the verification effect might have been exaggerated and the discounting effect underestimated.

A third type of limitation concerns the nature of the negative beliefs that comprised our statements—impersonal and somewhat abstract instead of personal and idiosyncratic. Given our approach of aligning design with paradigms of misinformation correction and process-dissociation, it was not possible to focus on personal beliefs. It would be difficult to imagine that summaries of research findings could influence the therapy client's beliefs about his or her individual situation. Thus, the application of this research to any sort of change beyond the more abstract beliefs of depressed people awaits evidence. Kube and Rozenkrantz's (2021) recent framework for understanding the processes involved in stages of updating applies to various measures of functional and dysfunctional and mostly idiosyncratic beliefs. To augment that framework, we suggest that both habitual rumination and

recollection of belief-relevant challenges will also emerge as important aspects of success and failure of attempts to modify idiosyncratic beliefs, if appropriate methods can be developed.

Relation to Other Investigations of Clinically Relevant Belief Change

Surprisingly little research has attempted to modify depressive beliefs of an impersonal nature.[3] On the other hand, there is a burgeoning literature that demonstrates the cognitive and behavioral inflexibility of depressed people in a variety of tasks that reveal beliefs more indirectly (see the reviews by Kube & Rozenkrantz, 2021; Stange et al., 2017). Recent evidence concerning two specific types of inflexibility are notable in relation to our results. First, beliefs become expectations when they refer to what might happen specifically to oneself in the future. In this regard, we note research concerning the difficulty of modifying negative performance expectations held by depressed people, even after receiving feedback that disconfirms their expectations (Kube, Glombiewski et al., 2019; Kube, Rief et al., 2019). These researchers coined the term "cognitive immunization" to represent strategies—such as the reappraisal of positive feedback—for the maintenance of negative expectations in the face of disconfirming evidence. Such inflexibility indicates motivational factors (i.e. reappraisal) that should exacerbate habit and counter any effects of remembering the evidence.

The second relevant line of research is a subset of experiments designed to change the negativity of depressive interpretation biases through cognitive bias modification—procedures that train non-negative resolutions of ambiguity inherent in descriptions of everyday situations (see LeMoult & Gotlib, 2019). Results from this paradigm occasionally have suggested inflexibility in depression (e.g., poor transfer to new tasks; LeMoult et al., 2018). Recently, however, the issue of inflexibility was addressed

[3] One somewhat irrational yet common belief—that not washing your hands after using the toilet will make you ill—was effectively changed through the use of "the behavioral experiment," a technique from cognitive behavioral therapy (McManus et al., 2011). As the name suggests, this procedure involves the consideration of evidence obtained through the design and simple test of an experiment performed by the client during a therapeutic session. Compared to a no-treatment control, this approach to establishing disconfirming evidence was successful in the nonselective sample.

more directly by Everaert et al. (2018), with a procedure to train revision of negative interpretations on the basis of disconfirming new information; participants with depression (or with social anxiety) resisted such revision. Our results suggest future explorations of whether strong interpretative habits, impaired memory for the new information, or both directly underlie resistance to the revision of negative interpretations (see Hertel et al., 2014). Such reduced flexibility, tapped lightly in our student sample, has the potential for very serious consequences. For example, Everaert et al. (2021) found that resistance to revision of negative interpretation predicts suicidal thoughts, as mediated by beliefs about being a burden to others.

We conclude with the observation that resistance to the modification of beliefs, expectancies, and interpretations in depression, although amply documented, is only beginning to be understood in terms of the processes that contribute to the phenomenon. Our results emphasize the strength of habitual aspects of belief. Our results are also consistent with the literature on factors establishing resistance during the anticipation, interpretation, and appraisal of new information (see Kube & Rozenkrantz, 2021). And they take a step toward understanding the contributions of memorial deficit or bias.

Acknowledgements

The authors are grateful for the dedicated help from Nabeel Gaber (especially), Rebecca Sanchez, and Emily Warkentin. These experiments were conducted according to the guidelines for the ethical treatment of human participants, and approval was obtained by the university's Institutional Review Board.

Online Supplement

Supplementary material related to this article can be found, in the online version, at [insert here]/

.

References

Abramson, L. Y., Metalsky, G. I., & Alloy, L. B. (1989). Hopelessness depression: A theory-based subtype of depression. *Psychological Review, 96*, 358-372.

Beck, A. T. (1967). *Depression: Clinical, Experimental, and theoretical aspects*. New York: Hoeber Medical Division, Harper & Row.

Beck, A. T., Steer R. A., & Brown, G. K. (1996). Manual for the Beck Depression Inventory (2nd Ed.). San Antonio, TX: Psychological Corporation.

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General, 121*, 446-458.

Everaert, J., Bronstein, M. V., Cannon, T. D., & Joormann, J. (2018). Looking through tinted glasses: Depression and social anxiety are related to both interpretation biases and inflexible negative interpretations. *Clinical Psychological Science, 6,* 517-528.

Everaert, J., Bronstein, M. V., Cannon, T. D., Klonsky, E. D., & Joormann, J. (2021). Inflexible interpretations of ambiguous social situations: A novel predictor of suicidal ideation and the beliefs that inspire it. *Clinical Psychological Science,* doi.org/10.1177/2167702621993867

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: Crrent status and future directions. *Annual Review of Clinical Psychology, 6,* 285-312.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior, 16,* 107–112.

Hertel, P. T. (2004). Memory for emotional and nonemotional events in depression: A question of habit? In D. Reisberg & P. Hertel (Eds.), *Memory and emotion (pp. 186-216)*. New York: Oxford University Press.

Hertel, P. T. Holmes, M., & Benbow, A. (2014). Interpretive habit is strengthened by cognitive bias modification. *Memory*, *22*: 737-746.

Hertel, P. T., & Rude, S. S. (1991). Depressive deficits in memory: Focusing attention improves subsequent recall. *Journal of Experimental Psychology: General, 120,* 301-309.

Hinze, S. R., Slaten, D. G., Horton, W. S., Jenkins, R., & Rapp, D. N. (2014). Pilgrims sailing the Titanic: Plausibility effects on memory for facts and errors. *Memory & Cognition*, *42*, 305–324.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541.

Jacoby, L. L., Baker, J. G., & Brooks, L R. (1989). Episodic effects on picture identification: Implications for theories of concept learning and theories of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 686-700.

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1420-1436.

Kube, T., Glombiewski, J. A., Gall, J., Toussaint, L., Gärtner, T., & Rief, W. (2019). How to modify persisting negative expectations in major depression? An experimental study comparing three strategies to inhibit cognitive immunization against novel positive experiences. *Journal of Affective Disorders*, *250C*, 231-240.

Kube, T., Rief, W., Gollwitzer, M., Gärtner, T., & Glombiewski, J. A. (2019). Why dysfunctional expectations in depression persist—results from two experimental studies investigating cognitive immunization. *Psychological Medicine, 49,* 1532–1544.

Kube, T., & Rozenkrantz, L. (2021). When beliefs face reality: An integrative review of belief updating in mental health and illness. *Perspectives on Psychological Science*, *16*(2), 247-274.

LeMoult, J., Colich, N., Joormann, J., Singh, M. K., Eggleston, C., & Gotlib, I. H. (2018). Interpretation bias training in depressed adolescents: Near- and far-transfer effects. *Journal of Abnormal Child Psychology, 46,* 159-167.

LeMoult, J., & Gotlib, I. H. (2019). Depression: A cognitive perspective. *Clinical Psychology Review, 69,* 51-66.

Lewandowsky, S., Ecker, U. K H., & Cook J. (2017). Beyond misinformation: Unerstanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition., 6,* 353-369.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13,* 106–131.

McManus, F., Van Doorn, K., & Yiend, J. (2011). Examining the effects of thought records and behavioral experiments in instigating belief change. *Journal of Behavior Therapy and Experimental Psychiatry, 43,* 540-547.

Rapp, D. N. (2016). The consequences of reading inaccurate information. *Current Directions in Psychological Science, 25*: 281-285.

Sharot, T., & Garrett, N. (2016). Forming beliefs: Why valence matters. *Trends in Cognitive Sciences, 20,* 25–33.

Stange, J. P., Alloy, L. B., & Fresco, D. M. (2017). Inflexibility as a vulnerability to depression: A systematic qualitative review. *Clinical Psychology: Science and Practice, 24,* 245–276.

Van Boekel, M., Lassonde, K. A., O'Brien, E. J., & Kendeou P. (2017). Source credibility and the processing of refutation texts. *Memory & Cognition, 45,* 168-181.

Wahlheim, C. N., Alexander, T. R., & Peske, C. D. (2020). Reminders of everyday misinformation statements can enhance memory for and beliefs in corrections of those statements in the short term. *Psychological Science, 31,* 1325-1339.

Watkins, E. R., & Nolen-Hoeksema, S. (2014). A habit-goal framework of depressive rumination. *Journal of Abnormal Psychology, 123,* 24–34.

Appendix

Background Script for the Session-1 Task

During this session, the monitor will display a series of statements that represent common beliefs. We know these beliefs are common because we consulted a research article listing the most widely held beliefs by individuals in a large random sample. We then conducted our own survey for replication and found similar results in the local population. We also investigated whether there was any empirical basis for these beliefs. So, after each belief statement, we have included a short summary of a research article that we found in our investigation that relates to the belief. These articles were published in reputable sources, and the findings have been replicated by other scientists. We also include a citation of the article which, in psychology, is the author's surname and the date of publication. If any belief really sticks out to you, please let me know at the end of the session, and I can send you the related article after the second session. Do you have any questions so far about the common beliefs or research articles?

As you will see, a few of the belief statements are fairly broad and general in nature, so they cannot be directly tested. However, the researchers involved in the relevant studies found ways to operationalize the terms in the statements, allowing us to generalize specific, measurable phenomena to broader claims. For example, if you were to see a statement like, "people work hard to achieve their goals," it's true that people's goals vary greatly, and the definition of "hard work" might be somewhat relative. But a study could still be conducted by figuring out how to measure variables such as degrees of desire for achieving goals, amount of time spent working toward goals, and general work habits. With appropriate sampling, the findings could then be extended to a broader population. Do you have any questions about how these beliefs have been operationalized?

Table 1

Mean Scores on the BDI and Ratings in Response to Debriefing questions (SD in Parentheses)

	Experiment 1		Experiment 2	
	Others (*n* = 24)	**Dysphoric (*n* = 24)**	**Others (*n* = 45)**	**Dysphoric (*n* = 45)**
Debriefing question				
Were you able to fully understand the belief and its related evidence, before the program moved on to the next one?	5.5 (1.02)	5.4 (1.38)	5.3 (1.26)	4.9 (1.44)
Did you have trouble remembering the belief while examining the evidence?	4.2 (1.88)	4.8 (1.81)	4.4 (1.68)	4.8 (1.43)
Did you feel discomfort from the emotional nature of some beliefs?	1.7 (0.96)	2.0 (1.43)	1.5 (1.22)	2.0 (1.36)*
Were the beliefs and related evidence presented relevant to you?	4.1 (1.38)	5.2 (1.37)*	4.1 (1.69)	5.0 (1.17)*
Did the beliefs seem like ideas some people would actually endorse?	6.2 (0.88)	6.1 (0.90)	6.4 (0.66)	6.0 (1.14)*
Did the evidence seem plausible?	5.8 (0.83)	5.5 (1.10)	5.8 (1.17)	5.3 (1.16)*
BDI-II score (in Session 1)	3.5 (2.84)	27.5 (8.03)	4.3 (2.39)	24.8 (7.85)
BDI-II score (in Session 2)	2.5 (2.55)	25.8 (8.32)	3.8 (2.77)	23.5 (8.13)

Note. The rating scale for the survey items ranged from ranged from 1 (“not at all”) to 7 (“very much so”). Means were not significantly different between the two groups unless the entry for the dysphoric group is followed by an asterisk. In that case a two-tailed t test revealed a difference, $p < .05$. BDI-II differences were not statistically evaluated, given their basis for selection.

Table 2

Experiment 1: Mean Accuracy Judgments in Session 1 and Belief Ratings in Session 2 (Standard Deviations in Parentheses)

	Evidence type			
Measure	**Verify**	**Discount**	**New statements**	**Average**
Percentage judged accurately in Session 1				
Dysphoric students	84.7 (12.92)	85.4 (15.77)	---	85.1 (11.77)
Others	91.0 (13.88)	86.8 (13.88)	---	88.9 (10.90)
Average	87.8 (13.63)	86.1 (14.71)	---	
Belief rating in Session 2				
Dysphoric students	5.7 (0.73)	4.4 (1.25)	4.9 (1.00)	5.0 (.619)
Others	5.3 (0.76)	3.3 (0.95)	4.1 (0.80)	4.3 (.572)
Average	5.5 (0.76)	3.9 (1.22)	4.5 (0.98)	
Proportion believed in Session 2				
Dysphoric students	.76 (.161)	.50 (.307)	.60 (.170)	.62 (.127)
Others	.69 (.201)	.28 (.217)	.46 (.216)	.48 (.163)
Average	.73 (.183)	.39 (.285)	.53 (.204)	

Note. The rating scale ranged from "strongly disbelieve" (1) and "strongly believe" (8). Mean belief ratings in Session 2 were contingent on accuracy in Session 1. Proportions believed were obtained by dichotomizing the ratings.