Turning quickly on myself: Automatic interpretation biases in dysphoria are self-referent

Alexandra H. Cowden Hindash & Jonathan Rottenberg

To cite this article: Alexandra H. Cowden Hindash & Jonathan Rottenberg (2017) Turning quickly on myself: Automatic interpretation biases in dysphoria are self-referent, Cognition and Emotion, 31:2, 395-402, DOI: 10.1080/02699931.2015.1105792

To link to this article: https://doi.org/10.1080/02699931.2015.1105792

Published online: 03 Nov 2015.
BRIEF ARTICLE

Turning quickly on myself: Automatic interpretation biases in dysphoria are self-referent

Alexandra H. Cowden Hindash and Jonathan Rottenberg

Psychology Department, University of South Florida, Tampa, FL, USA

ABSTRACT
Cognitive theories emphasise automatic interpretation biases (AIB) in the development and maintenance of depression. The current study examined AIB using the word sentence association paradigm for depression (WSAP-D) via endorsement rates and reaction time indices. We directly tested the importance of self-relevance for AIB by modifying the WASP-D task to include self-referent ambiguous stimuli and contrasting them with other-referent ambiguous stimuli. We hypothesised that the dysphoric group, but not the non-dysphoric group, would demonstrate AIB only for self-referent ambiguous stimuli. Consistent with our main hypotheses, dysphoric individuals endorsed negative interpretations more often and faster than non-dysphoric individuals, only for self-referent ambiguous stimuli. Self-relevance may be a critical aspect of AIB in dysphoric populations.

Depression is a debilitating disorder, with increasing prevalence and societal burdens (Kessler et al., 2003). The high costs associated with depression have motivated research into depression-associated risk factors. Cognitive theories of depression emphasise information processing factors in depression’s etiology and maintenance (Beck & Haigh, 2014; Williams, Watts, MacLeod, & Mathews, 1988). Specifically, depression has been associated with negative information processing biases, including biases in memory, attention, and interpretation (for review see Gotlib & Joormann, 2010).

Though relatively understudied, negative biases in interpretation are central to cognitive theories of depression (Beck & Haigh, 2014). Interpretation is the information processing component that assigns meaning to environmental information. Interpretations integrate information from attention and memory to make overall sense of current and past environmental information (Beck & Haigh, 2014). Interpretations can be accomplished via both automatic and controlled cognitive processes, which likely exist on a continuum and interact with one another (Beevers, 2005).

At the automatic level, information is interpreted quickly by integrating what is available in the environment within the first few seconds of stimulus presentation. Although there is no consensus about the exact boundaries of an automatic interpretation, a strict definition would denote interpretations that are formed quickly without any conscious effort to assign meaning to an ambiguous stimulus. Automatic interpretations have been historically difficult to record because of the timescale at which they occur. By contrast, information processing at the controlled level occurs over several seconds or even minutes and allows further analysis, elaboration, and integration with other current and past environmental information to assign meaning to an ambiguous stimulus (Beevers, 2005). Controlled, elaborative interpretations have been relatively easier to record because the process occurs more slowly, and is therefore easier to observe than automatic interpretations. Metrics that span the continuum from automatic to controlled interpretation over milliseconds to minutes, range from reflexive, physiological responses such as startle (Lawson, MacLeod, & Hammond, 2002), to speeded decision reaction times (Cowden Hindash & Amir, 2012), to

© 2015 Taylor & Francis
CONTACT Jonathan Rottenberg rottenberg@usf.edu
http://dx.doi.org/10.1080/02699931.2015.1105792

**Elaborative interpretation biases in depression**

Historically, much work has assessed interpretation biases in depression via self-report (for reviews see Mathews & MacLeod, 2005; Wisco, 2009). Self-report interpretation bias measures generally present ambiguous scenarios followed by potential explanations of those scenarios, and participants must endorse one potential explanation as most likely. Although self-report measures can provide some insight into interpretation biases, self-report questionnaires are also prone to response, expectancy, and comparative biases. Such biases enter in when respondents are given time to deliberate about all presented experimental scenarios prior to endorsement. Because of the amount of deliberation involved, these measures can only assess *elaborative* interpretation biases rather than *automatic* interpretation biases (AIB).

To reduce the influence of response and expectancy biases, Wisco and Nolen-Hoeksema (2010) devised an elaborative interpretation bias measure called the Interpretation Bias Questionnaire (IBQ). The IBQ asks individuals to self-generate interpretations of ambiguous scenarios. For each scenario, participants must self-generate at least four interpretations and then indicate which interpretation they think is most likely to occur. The scenarios are either self- or other-referent and address social or achievement-based themes. Using the IBQ, Wisco and Nolen-Hoeksema (2010) found that, compared to healthy controls, depressed and dysphoric individuals were more likely to generate negative interpretations of ambiguous scenarios and judged negative interpretations as most likely to occur. Importantly, a negative interpretation bias occurred only when scenarios were self-referent. For other-referent scenarios, dysphoric and depressed individuals were more likely than non-dysphoric controls to endorse neutral or positive scenarios as most likely. This suggests elaborative interpretation biases for negative material may be specific to self-relevant material.

**AIB in depression**

Despite theory that posits depression-related interpretation biases both for automatic *and* elaborative cognitive processing (Beevers, 2005), AIB have been difficult to detect in experiments. While experimental priming paradigms have found AIB in other psychiatric disorders (for review see Mathews & MacLeod, 2005), there have been several null findings in dysphoric or depressed samples (Bisson & Sears, 2007; Lawson & MacLeod, 1999; Mogg, Bradbury, & Bradley, 2006). These null results are doubly unexpected given evidence that depressed individuals negatively interpret simple, single word neutral cues (Siegle, Granholm, Ingram, & Matt, 2001, 2003), homophones (Mogg et al., 2006), and scrambled sentences under cognitive load (Wenzlaff & Bates, 1998). Wisco (2009) concluded that the lack of evidence of AIB might mean that negative interpretation biases function only at extended, elaborative processing levels in depression.

More recently, however, experiments using semantic association paradigms provided evidence of automatic semantic negative AIB in a dysphoric sample (Cowden Hindash & Amir, 2012; Sears, Suzie Bisson, & Nielsen, 2011). These semantic association paradigms present ambiguous sentences followed by unambiguous related words. Participants indicate as quickly as possible whether or not (via yes/no button push) the word is related to the sentence. These paradigms allow for the measurement of two aspects of AIB via aggregate indices of endorsement rates (i.e., how often an individual indicates a word type is related to the sentence) and trial-by-trial response times (i.e., how quickly the individual indicates the word is related to the sentence). While Cowden Hindash and Amir (2012) found automatic negative AIB in a dysphoric sample with both endorsement rates and response times, Sears et al. (2011) found evidence of automatic negative AIB only in endorsement rates, which debatably are relatively less automatic. Interestingly, both of these AIB studies featured self-referent ambiguous stimuli, which suggests the possibility that previous null findings using complex semantic stimuli (Bisson & Sears, 2007; Lawson & MacLeod, 1999; Mogg et al., 2006) may reflect the use of other-referent stimuli.

**Centrality of self-relevance**

The centrality of negative self-focus in depressive mood states (i.e., negative self-perceptions and negative self-attributions; Haaga, Dyck, & Ernst, 1991) may help explain why other-referent stimuli from semantic information paradigms do not produce evidence of
AIB in depressed or dysphoric samples (Bisson & Sears, 2007; Lawson & MacLeod, 1999; Mogg et al., 2006). Although intriguing, this explanation is speculative because studies using self-referent stimuli (Cowden Hindash & Amir, 2012; Sears et al., 2011; Wenzlaff & Bates, 1998) have not included conditions with other-referent stimuli, which would allow comparison of AIB between self- and other-referent stimuli.

Current study

The current study used a semantic association paradigm (Cowden Hindash & Amir, 2012) and directly examined the role of self-relevance in AIB. We used ambiguous stimuli that were either self- or other-referent and conducted between-group comparisons on both the referent of the ambiguous stimulus (sentence) and the valence (word) of the interpretations. To strengthen confidence in the results, we indexed AIB via two kinds of metrics (endorsement rates and response times). We hypothesised that dysphoric individuals would demonstrate negative AIB with self-referent ambiguous stimuli. By contrast, we expected no AIB for other-referent ambiguous stimuli.

Method

Participants

One hundred and sixty-three participants were recruited through the psychology research subject pool at a large university. All received course credit for participation. Of the 163 participants who provided informed consent, 48 were excluded from analyses for either computerised task non-completion (computer failure, time restriction n = 6) or for falling outside standard dysphoria cut-offs (n = 42; Sprinkle et al., 2002). Dysphoric and non-dysphoric groups were formed according to standard cut-offs (Sprinkle et al., 2002), based on Beck Depression Inventory—Second Edition scores (BDI-II; Beck, Steer, & Brown, 1996). The dysphoric group (n = 42) were individuals who scored a 14 or higher on the BDI-II. The non-dysphoric group (n = 73) were individuals who scored a 7 or lower on the BDI-II. Dysphoric and non-dysphoric groups did not differ on age (20.17 vs. 22.55; p = .22), gender (76.2% vs. 64.4% female; p = .19), education (9.6% vs. 9.5% in last year of college), or ethnic background (30.1% vs. 38.1% Caucasian; p = .19). As expected, relative to the non-dysphoric group, the dysphoric group self-reported higher levels of depressive symptoms (M = 20.98 (5.4) vs. M = 3.58 (2.4), t(113) = 23.56, p < .001), trait anxiety (M = 56.31 (8.1) vs. M = 32.05 (7.6), t(113) = 16.09, p < .001), and state anxiety (M = 47.90 (12.9) vs. M = 29.92 (7.4), t(113) = 9.51, p < .001).

Materials

Self-report measures

To assess psychiatric symptoms and affect, participants completed the BDI-II (Beck et al., 1996), the Spielberger State/Trait Anxiety Inventory (STAI-S/T; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983), and the Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988). In the current sample, internal consistency of these self-report measures was good (α = .93 and .95 for BDI-II and STAI-S/T, respectively; PANAS negative affect, α = .84, PANAS positive affect, α = .93).

Computerised bias assessment

To examine AIB, we used the Word Sentence Association Paradigm for Depression (WSAP-D; Cowden Hindash & Amir, 2012), a computerised semantic association task presented in E-Prime 2.0. Participants were seated approximately 52 cm from the monitor. Stimuli were presented in black Times New Roman 14pt font on a white background. Each WSAP-D trial began with a fixation cross centred on the screen. Participants were instructed to focus on the cross, which was replaced by one of 170 sentences. The sentence was replaced by a single paired unambiguous word (either negative or benign) after 1000 ms. Sentences were presented for only 1000 ms to curtail elaboration on the material. Upon seeing a word, participants then judged whether the word was related to the sentence. Participants were instructed to make this judgement as quickly as possible via a mouse click. Judgement reaction times assessed perceived relatedness of the semantic material. A “yes” response constituted endorsement of a relationship between the word and the sentence, while a “no” response constituted rejection of a relationship between the word and the sentence. Prior to the main task, participants completed practice trials with unambiguous sentences and obviously related or unrelated words. Participants’ average accuracy rate was 97% on practice trials once participants learned the correct mouse button to push (i.e., most errors occurred during trials 1–4).

To isolate the role of self-relevance, the ambiguous sentences took self-referent or other-referent forms. In
the self-referent form, the subject was “you,” whereas in the other-referent form the subject was another named person (e.g., People tell you/Katie to smile more often). Paired unambiguous words remained the same regardless of the sentence referent. Each participant was presented with 85 self-referent sentences and 85 other-referent sentences, with 43 negative words and 42 benign words for each referent type. All sentences and conditions were randomly ordered without presentation restriction for each participant. That is, each participant viewed an individualised randomly ordered list of the 170 base sentences paired with one of the two possible unambiguous words for each sentence.

The WSAP-D yields two measures of AIB: endorsement rates (i.e., how often the participant answered “yes”) and response times (i.e., how quickly the participant decided). Both indices are considered automatic because both are based on a single speeded behavioural response in a context where participants are given minimal time to elaborate on information presented in an ambiguous sentence. Endorsement rates were the aggregate proportion of a word type (benign/negative) endorsed as related to the sentence out of all possible endorsements. For example, the negative self-referent endorsement rate is calculated by dividing the number of trials where the negative self-referent interpretation was endorsed (i.e., the participant answered “yes”) by the total number (i.e., 43) of negative self-referent trials. A zero would indicate that a relationship between the word type and sentence was never endorsed, while a one would indicate that a participant always endorsed a relationship between the word type and sentence.

The response time indices were the individual mean reaction times for each of eight combinations of word valence, response, and sentence referent (e.g., endorsing negative self-referent interpretations, rejecting benign other-referent interpretations). Faster response times to endorse a sentence–word relationship imply a participant had automatically interpreted the ambiguous sentence in line with the valence of the unambiguous word (i.e., requiring less time to process the information). For example, a faster response time to endorse rather than reject negative words as related to self-referent sentences would indicate ready interpretation of the sentence as negative (i.e., less time was required to connect the meaning of the word in relationship to the sentence). Prior to analysis, WSAP-D trials with reaction times less than 200 ms, greater than 5000 ms, or responses greater than ± 2.5 standard deviations from the mean response latency for each participant were defined as outliers and excluded from all analyses. This resulted in exclusion of 5.63% of trials. We did not restrict reaction times to a specific cut-off (e.g., 1500 ms) to allow for individual differences in processing speed. For group comparisons, individual means for each trial type were averaged to form group means for each trial type.

Procedure
Participants provided informed consent upon arrival to the laboratory. Participants then completed the self-report measure packet followed by one of two computer tasks, the WSAP-D and a computerised version of the IBQ (Wisco & Nolen-Hoeksema, 2010; data not reported here). After completing the first computer task, participants completed the scrambled sentences task (SST; Wenzlaff & Bates, 1998) and a stressful anagrams task, followed by the second computer task. The order of computer tasks was counterbalanced, so that half of participants completed the WSAP-D before the SST and the other half completed the WSAP-D after the SST. Participants completed a PANAS rating after each task. Task order had no effect on any analysis and was not considered further.

Results
Endorsement rates
We expected that dysphoric individuals would endorse negative interpretations of ambiguous sentences when ambiguous sentences were self-referent. In our design, we addressed this question with a 2 (Referent: self vs. other) X 2 (Valence: negative vs. benign) X 2 (Group: dysphoric vs. non-dysphoric) ANOVA, with repeated measurement on the first two factors. In our design, thus, we expected to observe a three-way referent*valence*group interaction. Our omnibus analysis revealed a main effect of Valence \( F(1,113) = 207.71, \ p < .001, \ \eta^2 = .65 \), modified by two-way interactions between Valence and Group \( F(1,72) = 6.54, \ p = .012, \ \eta^2 = .06 \), and between Referent and Group \( F(1,113) = 5.01, \ p = .027, \ \eta^2 = .04 \). More importantly, these effects were modified by the expected three-way interaction between referent, word valence, and group \( F(1,113) = 4.89, \ p = .029, \ \eta^2 = .04 \).

To better understand this three-way interaction, we examined the two-way interaction between group and valence, separately for each referent condition.
In the other-referent condition, only the main effect of word valence was significant $F(1,113) = 172.47, p < .001, \eta^2 = .60$, such that individuals from both groups were more likely to endorse benign interpretations than negative interpretations. By contrast, in the self-referent condition, we observed a main effect of word valence $F(1,113) = 193.05, p < .001, \eta^2 = .63$, modified by a group by word valence interaction $F(1,113) = 10.81, p = .001, \eta^2 = .09$. Consistent with our hypotheses, follow-up independent samples t-tests revealed that for self-referent ambiguous sentences dysphoric individuals had a higher endorsement rate for negative interpretations ($M = 50.13, SD = .15$) than non-dysphoric individuals ($M = 40.9, SD = .17$), $t(113) = -2.900, p < .01, d = .55$ (Figure 1). By contrast, endorsement rates of benign interpretations of self-referent ambiguous sentences did not differ by group. These results were robust to covariate analyses which included trait and state anxiety and positive and negative affect. Both anxiety, assessed by the STAI-S/T, and affect, assessed by pre-task PANAS, did not influence results or act as significant factors in interpretation biases.

**Reaction times**

In reaction time analyses, we expected the dysphoric group to be slower to endorse benign words and faster to reject benign words as related to a self-referent (but not other-referent) ambiguous sentence. We examined reaction time indices of interpretation biases with a 2 (Referent: self vs. other) X 2 (Valence: negative vs. benign) X 2 (Response: endorse vs. reject) X 2 (Group: dysphoric vs. non-dysphoric) ANOVA with repeated measurement on the first three factors. From this omnibus four-way interaction, we break the analyses into progressively simpler effects by examining lower-level interactions, especially those that involved group differences. These analyses, first, revealed a significant omnibus four-way interaction between Referent, Valence, Response, and Group $F(1,113) = 7.50, p = .007, \eta^2 = .06$. Consistent with our expectations, this overall model suggests that the groups differ in their response times depending on the valence of the word and the referent of the sentence.

To gain a better understanding of this four-way interaction, the three-way interaction between Referent, Valence, and Group was examined separately by Response in order to test for a relationship between word valence, group, and sentence referent. We broke effects down by response type (endorse/reject) to address our primary hypothesis: that groups differ in the valence of their interpretation biases, but only when ambiguous sentences are self-referent. In the endorsement conditions, only main effects were significant. However, in the rejection conditions, there was a three-way interaction between
Referent, Valence, and Group $F(1,113) = 8.08, p = .005, \eta^2 = .07$. This indicates that the dysphoric and non-dysphoric groups differ from each other when rejecting interpretations but not when endorsing interpretations.

To better understand this three-way interaction in the rejection conditions, we tested the two-way interaction between valence and group separately by referent conditions. As with our endorsement rate analyses, the valence by group interaction was not significant for other-referent sentences; nor were any other effects. By contrast, for self-referent sentences, we observed a Group by Valence interaction $F(1,113) = 15.99, p < .001, \eta^2 = .12$. This indicates that groups differ in how quickly they reject a relationship between the word and the sentence based on the valence of the word. Follow-up independent t-tests revealed significant group differences on trials rejecting self-referent sentences with benign words $t(113) = 3.09, p < .01, d = .58$ such that dysphoric individuals ($M = 1309.67$, $SD = 467.67$) were faster to reject benign interpretations of self-referent ambiguous sentences than non-dysphoric individuals ($M = 1600.92$, $SD = 497.61$). Independent t-tests also revealed group differences in time to endorse self-referent sentences with negative words $t(113) = 2.08, p = .04, d = .39$, such that dysphoric individuals ($M = 1251.45$, $SD = 446.40$) were faster to endorse negative interpretations than non-dysphoric individuals ($M = 1421.01$, $SD = 418.48$), although these differences should be interpreted with caution as the interaction effects missed conventional levels of significance ($p = .08$). Group differences in reaction times in each trial type are shown in Figure 2. In sum, dysphoric individuals were faster to both endorse negative interpretations and to reject benign interpretations of ambiguous information, demonstrating negative AIB, but only when the ambiguous information was self-referent. As before, these group effects were robust to the addition of anxiety and affect covariates.

**Discussion**

Cognitive biases are established risk factors for depression and potentially provide important targets for intervention. Here, our purpose was to better characterise a relatively understudied cognitive bias in depression, automatic negative interpretation bias (AIB). More specifically, we directly examined how dysphoric persons assign meaning to ambiguous self-relevant information. Our main goal was to expand upon previous research which evidenced AIB in dysphoria (Cowden Hindash & Amir, 2012; Sears et al., 2011), and suggested the possibility that AIB hinges upon the self-relevancy of the ambiguous information (Cowden Hindash & Amir, 2012; Wisco, 2009). Here, we tested this idea directly, hypothesising that relative to non-dysphoric persons, dysphoric persons would demonstrate negative AIB only for self-referent material.
Data largely conformed to expectation: Across measures, the dysphoric group differed from the non-dysphoric group on indices of negative AIB only when ambiguous information was self-relevant. In the sections below, we consider the ways the study supports stimulus self-relevance as an important factor in shaping whether AIB is observed in depression (Cowden Hindash & Amir, 2012; Wisco, 2009).

Specifically, when sentences were self-relevant, the dysphoric group was more likely to indicate a relationship between a negative word and an ambiguous sentence than the non-dysphoric group, a pattern suggesting that dysphoric persons make negative self-associations more often than non-dysphoric persons. Further, the groups did not differ in endorsements of a relationship between a benign word and an ambiguous sentence. The reaction time data, like the endorsement rate results, further suggest depression may be characterised by negative AIB. Specifically, the dysphoric group was faster both to endorse the relationship between a negative word and an ambiguous sentence and to reject a relationship between a benign word and an ambiguous sentence. These differences suggest that dysphoric persons are able to make faster decisions about whether or not the word is related to the sentence because they are assigning negative meaning to the ambiguous material more quickly than non-dysphoric persons. In other words, our findings suggest that dysphoric and non-dysphoric persons differ both in how often and how quickly they assign negative meaning to ambiguous information.

While these results are consistent with our hypothesis, our design was complex, and we acknowledge that other patterns within the reaction time indices could also indicate negative AIB. For example, an alternate reaction time pattern broadly consistent with our hypothesis would be if, compared to the non-dysphoric group, the dysphoric group more slowly rejected a relationship between the ambiguous sentence and a negative word. Nevertheless, we focused on an interaction pattern to allow continuity with earlier research (Beard & Amir, 2009; Cowden Hindash & Amir, 2012). Notably, the patterning of negative AIB observed here with reaction times was consistent with those observed with our other assessment index (endorsement rates) and with previous studies (Cowden Hindash & Amir, 2012; Sears et al., 2011).

Importantly, within our design, we are able to interpret these interpretations as automatic because of the quick processing required to complete the WSAP-D. The WSAP-D allowed participants only one second to process the ambiguous sentence before they must indicate whether or not the word is related to the sentence. Although there are no formal criteria for what constitutes automatic processing, and we cannot rule out that some degree of controlled processing occurred, our test context required participants to make rapid judgements without full confidence. In debriefing, participants often reported feeling uncertain or anxious about their decisions, spontaneously volunteering to experimenters that they were barely able to read the sentence before it was replaced by the word. Despite uncertainty about their endorsement decisions, high accuracy rates observed during the practice trials suggest participants could process key features of the sentences with limited conscious awareness.

There are some limitations to the present study, which should be addressed in future experiments. First, our sample of dysphoric students was not diagnosed and thus generalisability to clinical depression is uncertain (Lewinsohn, Solomon, Seeley, & Zeiss, 2000; Vredenburg, Flett, & Krames, 1993). Second, while we directly contrasted the effects of self-referent versus other-referent ambiguous stimuli, the other-referent stimuli referred to unknown and random individuals. It has been postulated that the nature of interpretation biases may change, at least for elaborative processing, when other-referent material refers to known persons (Wisco & Nolen-Hoeksema, 2010), an idea that should be tested in future experiments. Third, while we focused on assessing AIB, the interrelation between automatic and elaborative interpretation biases is as yet poorly understood, so we cannot assume our sample would also demonstrate elaborative interpretation biases. Finally, although our measures have been accepted as reflecting AIB and are in keeping with other laboratories (Beard & Amir, 2009; Möbius, Tendolkar, Lohner, Baltussen, & Becker, 2015; Sears et al., 2011), automaticity exists on a continuum and we cannot preclude the possibility that some degree of elaboration took place during the WSAP-D task.

Despite these limitations, the present study possesses several notable strengths. First, we used an experimental paradigm to assess AIB in complex semantic stimuli, which have been historically difficult to assess and to find in dysphoric or depressed samples. Our paradigm, the WSAP-D, used two separable bias indices, both of which yielded evidence of negative AIB. Indeed, some of our findings replicated
previous work using semantic association paradigms to assess AIB (Cowden Hindash & Amir, 2012; Sears et al., 2011). Finally, we interpret these data as suggesting that self-relevance is critical in determining whether AIB is observed. We found negative AIB only for self-referent ambiguous information and not for other-referent material. This supports theoretical suggestions that depressive cognitive biases—at least in the area of interpretation—are deeply rooted in self-relevance (Wisco, 2009), a pattern that may help explain the repeated difficulties in observing AIB in depression over the last 15 years (Bisson & Sears, 2007; Lawson & MacLeod, 1999; Mogg et al., 2006).

Disclosure statement
No potential conflict of interest was reported by the authors.

ORCID
Alexandra H. Cowden Hindash http://orcid.org/0000-0001-6113-131X

References