Developing a novel assessment of interpretation flexibility
Deng, W.; Everaert, J.; Creighton, M.; Bronstein, M.; Cannon, T.D.; Joormann, J.

Published in:
Personality and Individual Differences

DOI:
10.1016/j.paid.2022.111548

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Tilburg University Research Portal

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Developing a novel assessment of interpretation flexibility: Reliability, validity and clinical implications

Wisteria Deng a,*, Jonas Everaert b, c, Mackenzie Creighton d, Michael V. Bronstein e, Tyrone Cannon a, Jutta Joormann a

Abstract
Interpretation bias and inflexibility have been implicated in a wide range of psychopathologies, including affective disorders and disorders involving persecutory ideation. Existing tasks that measure both interpretation bias and inflexibility (which is essential as bias can make individuals appear inflexible) rely heavily on verbal stimuli, which may be suboptimal for certain applications, including neuroimaging and developmental studies. To overcome these limitations, a picture-based task, the Interpretation Inflexibility Task (IIT) was developed to simultaneously assess interpretation bias and inflexibility. The present manuscript demonstrates the strong convergent validity of the IIT and an existing, verbal measure of interpretation bias and inflexibility (the Emotional BADE task), in a general population sample (N = 247). Across both tasks, inflexible negative interpretations (above and beyond interpretation bias) were associated with symptoms of depression, social anxiety, and paranoia. The consistency of this association across stimulus modalities suggests the existence of higher-level cognitive deficits that increase interpretation inflexibility and thereby increase risk for both affective and psychotic symptoms. The IIT provides an innovative and reliable paradigm to investigate this transdiagnostic factor, particularly given its superior suitability for neuroimaging studies.

1. Introduction
Aberrant interpretation of ambiguous information is a transdiagnostic factor that confers risk of experiencing several forms of psychopathology, including depression (Everaert, Podina, & Koster, 2017), social anxiety (Hirsch et al., 2020; Oglesby, Stentz, Portero, King, & Schmidt, 2019), and psychosis (Freeman et al., 2013; Lam, Mou-chliantitis, Lee, & Yiend, 2021; Trotta et al., 2021). Research suggests that these mental health conditions feature both biased and inflexible interpretations (Bronstein, Everaert, Castro, Joormann, & Cannon, 2019; Everaert, Bronstein, Cannon, & Joormann, 2018; Woodward, Moritz, Cuttler, & Whitman, 2006; Woodward, Moritz, Menon, & Klenge, 2008). Interpretation bias refers to the tendency to consistently resolve ambiguous situations in a characteristic direction – for example, as positive or negative; benign or threatening. Studies have shown that interpretation bias may be a contributing factor to the emergence and maintenance of affective (Chen, Short, & Kemps, 2020; Meixner, Montag, & Herbert, 2019) and psychotic disorders (Freeman et al., 2013; Hurley, Hodgkins, Coker, & Fowler, 2018). In addition, recent work suggests that the valence of interpretations may not be the most important factor that determines the impact on psychopathology. Rather, the ability to revise interpretations and adapt to evolving situations is crucial in understanding the cognitive mechanisms underlying symptoms (Mehu & Scherer, 2015). In particular, the concept of interpretation flexibility has gained attention in recent work on distorted interpretation processes in psychopathology. This concept refers to the ability to integrate new attributes of unfolding situations and update interpretations within changing circumstances (Everaert et al., 2018). Research indicates that interpretation inflexibility is associated with affective (Zhu, Kwok, Chan, Chan, & So, 2020) and psychotic symptoms across the severity continuum (Colbert, Peters, & Garety, 2016; Serrano-Guerrero, Ruiz-Veguilla, Martin-Rodriguez, & Rodriguez-Testal, 2020;
Further, modifying rigid thinking and revising biases in interpretation are well-established treatment goals for depression (Yasinski et al., 2020) and paranoid thinking (Geraets et al., 2020). The integral role of interpretation flexibility in understanding psychiatric symptoms and informing interventions indicates that it is not only important to consider what people believe (i.e., the content of emotional interpretations), but also how people revise those beliefs (i.e., in)flexibility in revising emotional interpretations) as ambiguous situations evolve over time.

To date, only one cognitive task, the emotional Bias Against Disconfirmatory Evidence (BADE), is designed to tease apart the effects of interpretation bias and interpretation inflexibility on psychopathology. The emotional BADE task presents social scenarios in gradual steps, prompting participants to revise their initial interpretations based on novel disconfirmatory information (Everaert et al., 2018). Specifically, participants are expected to revise their interpretations as subsequently encountered evidence disconfirms their initial impressions. The emotional BADE task builds on the original (non-emotional) BADE paradigm (Woodward et al., 2006) by providing new insights into how interpretations dynamically change as ambiguous self-referential situations gain clarity and shift in emotional valence over time. This task has consistently captured inflexibility and bias in interpretation processes across several studies (Everaert et al., 2018; Everaert, Bronstein, Cannon, Klonsky, & Joormann, 2021; Everaert, Bronstein, Castro, Cannon, & Joormann, 2020).

Whereas the existing emotional BADE task provides a measure for interpretation bias and inflexibility, it is limited by its reliance on verbal stimuli. Moreover, while initial versions of the original BADE task were picture-based (Woodward et al., 2006), this task was discontinued for several reasons (e.g., few scenarios were available; scenarios were presented in a less ecologically-valid, reverse temporal order). The IIT overcomes these limitations of the picture-based original BADE task, and builds on it by focusing on social scenarios and systematically manipulating the emotional valence of scenario resolutions. Developing a picture-based version of the emotional BADE task would be advantageous for several reasons. First, a picture-based task may be more externally valid. Ambiguous information is often presented in visual form as people navigate situations in their daily life (Hutmacher, 2019). In this format, cognitive deficits associated with psychopathology (e.g., in facial emotion recognition, theory of mind) could create additional ambiguity and hinder belief updating in ways that may be less apparent in tasks using verbal stimuli. A picture-based task may be a superior choice for investigating how interpretive processes are impacted by these deficits.

Second, visual stimuli, such as pictures, have been shown to be intrinsically more vivid, arousing and emotion-provoking than words (Bartoszek & Cervone, 2017; Glaser & Glaser, 1989). Especially when used in an emotional context, pictures produce a more consistent and stronger affective response from participants than text (Kensinger & Schacter, 2006). Indeed, pictures may better capture emotional valence and be particularly suitable for the measurement of interpretation bias and inflexibility in emotion-laden settings (Kensinger & Schacter, 2006). Further, a picture-based task (vs. a verbal one) may facilitate participants’ use of mental imagery to imagine themselves in the scenarios due to the greater similarity of pictures to perception (Holmes, Mathews, Mackintosh, & Dalgleish, 2008), allowing them to take on a more personal perspective as they are asked to interpret the scenarios (Fuchs et al., 2010). This mental imagery is critical because cognitive biases associated with psychopathologies featuring inflexible/biased interpretation, such as depression, are more apparent for self-referential stimuli (LeMoult, Kirkanski, Prasad, & Gotlib, 2017).

Moreover, the picture-based task also has advantages related to timing. With pictures instead of verbal stories as stimuli, overall test time may be reduced, promoting a more efficient implementation of the task. Further, a picture-based version may create fewer confounds related to individual differences in reading times. This provides a clear advantage for picture-based tasks in developmental studies (e.g., those examining trajectories of interpretation bias and inflexibility across adolescence) and in studies where it is critical to know when, exactly, a stimulus was processed (e.g., studies using neuroimaging to search for biobehavioral markers of interpretation bias and inflexibility).

A final reason for creating a picture-based task in the present study is to investigate whether the relations between inflexibility and symptoms of different psychopathologies (statistically controlling for interpretation bias) observed in previous studies using the text-based Emotional BADE Task persist when using a different stimulus modality. If the pattern of results is similar across modalities, this would implicate higher-level cognitive deficits (rather than deficits in any particular perceptual information-processing stream) as a transdiagnostic factor that contributes to affective and psychotic symptoms.

1.1. The present study

The present study details the development of the Interpretation Inflexibility Task (IIT), a novel, picture-based task designed to assess the process of evidence integration in gradual steps. In order to determine the convergent validity of the IIT and the Emotional BADE Task, we investigated whether these tasks produce a consistent pattern of associations between metrics of interpretation bias/inflexibility and symptoms of depression, social anxiety, and paranoia (all of which have been previously linked to inflexibility in studies using the Emotional BADE Task). We hypothesized that:

1) The IIT will exhibit good-to-excellent internal consistency.
2) Metrics of interpretation bias and inflexibility derived from the IIT and the Emotional BADE Task will be at least moderately correlated.
3) The pattern of results derived from both tasks will replicate those obtained in previous studies using the Emotional BADE Task. Specifically, we expected that inflexible negative interpretations in both tasks would be associated with greater severity of depression, social anxiety, and paranoia when statistically controlling for interpretation bias (inflexible negative, vs. positive, interpretations are the focus given their more consistent relation with symptoms in our previous studies). We also expected that positive interpretation bias would negatively correlate with symptoms of depression and social anxiety, whereas negative interpretation bias would be positively correlated with depressive symptoms and social anxiety.

2. Method

2.1. Participants

Participants (N = 274) were recruited via Amazon’s Mechanical Turk (MTurk), an online crowdsourcing platform that provides access to a large and diverse sample for mental health research studies (demographics: see Table 1). Participation in this study was restricted to MTurk users who were at least 18 years old and lived in the United States.

2.2. Data quality

Following recommendations for research using crowdsourced samples (Chandler & Shapiro, 2016), the study only recruited MTurk users who had a history of providing good-quality responses. Participants were required to have completed at least 500 MTurk studies and to have had their work approved (vs. rejected) in 98% of the studies they
completed previously on MTurk. In addition, three questions were included to discriminate attentive from inattentive participants. These questions were presented at random intervals and participants were required to answer all three correctly. Consistent with previous research, participants (n = 1) were excluded from all analyses if they completed the survey in <60% of the projected time (<27 min) (Everaert et al., 2018). Finally, duplicate IP addresses and suspicious geolocations were blocked and IP addresses were verified for consistency with a US location through CloudResearch (an online platform that provides additional services for researchers using MTurk). Research using a similar approach (e.g., requiring a history of good quality response) has demonstrated that MTurk data are comparable to those collected in the laboratory (Chandler & Shapiro, 2016).

2.3. Emotional BADE task

In the emotional BADE task (Everaert et al., 2018), participants were presented with 24 scenarios depicting common interpersonal situations. Each scenario consisted of three sentences, each revealing more information about the scenario. After each sentence of the scenario was presented, participants were asked to rate the plausibility of four interpretations of the information in the scenario, which were presented in randomized order. The rating scale used in this study was simplified to a range of “Poor” (a score of 1) to “Excellent” (a score of 7). For each part of every scenario, four interpretations were listed: one Absurd interpretation (which remained implausible throughout the scenario), two Lure interpretations (which were initially most plausible but became less plausible on the third sentence of the scenario), and one True interpretation (which was initially less plausible than the Lure interpretations but became the most plausible by the end of the scenario).

Two types of scenarios were included to examine whether interpretation inflexibility differed according to the valence of initial interpretations relative to that of the corresponding disconfirmatory evidence. The first scenario type, positive outcome scenarios, were designed to encourage initial negative interpretations, but by their end suggested that a more positive interpretation was warranted. For example, one scenario reads as follows: “The company you are working for needs to lay off many employees. You are called in to see your boss. You hear everyone starting to laugh,” “The other people think you have a great sense of humor”) and the True interpretation had a negative valence (e.g., “Some people think you can’t tell a joke properly”). The two scenario types were mixed and presented in randomized order across participants. While results from both scenario types are presented below, note that the metrics of interpretation bias/inflexibility derived from negative outcome scenarios have produced the most consistent association with symptoms in previous studies (Everaert et al., 2018).

2.4. Interpretation inflexibility task

The Interpretation Inflexibility Task (IIT) was developed by compiling pictures based on scenarios from the emotional BADE task (Everaert et al., 2018). The IIT follows the general structure of the emotional BADE Task. Each scenario gradually provides more information about an interpersonal situation. 12 scenarios have a positive resolution and the remaining 12 have a negative resolution. Instead of the verbal scenarios used in the Emotional BADE Task, each of the 24 IIT scenarios is based on a stock photo that is gradually revealed to respondents. To mirror the structure of the emotional BADE task in presenting each scenario in three parts, each IIT scenario presents the same stock photo three times: with 80%, 20%, and 0% of the photo blurred. The blurred regions were selected with the goal of obscuring the emotional valence of a given scenario. By gradually reducing the percentage of the photo that is blurred, the respondent receives more information that may help to resolve the initially ambiguous situation.

Similar to the emotional BADE task, two types of scenarios were included in the IIT to examine whether interpretation inflexibility differed according to the valence of the new evidence in relation to the initial interpretation that was most strongly suggested by ambiguous stimuli. Half of all IIT scenarios have a positive outcome (similar to that in the disconfirming-the-negative scenarios in the Emotional BADE Task). An example positive outcome scenario is shown in Fig. 1 (top). In this example, the three pictures (panels 1a-1c) progressively reveal more information about the scenario, culminating in panel 1c, which encourages a positive interpretation. An example negative outcome IIT scenario (analogous to the disconfirming-the-positive scenarios in the Emotional BADE Task) is shown in panels 2a-2c of Fig. 1. Each of the three pictures (panels 2a-2c) reveals more information about the scenario, culminating in panel 2c, which prompts a negative interpretation. Similar to the emotional BADE task, for each of the three pictures, four interpretations of the depicted scenario were presented to participants, who were asked to rate the plausibility of each interpretation. In contrast to the emotional BADE task (in which Lure statements were initially most likely), the first picture in each set of IIT scenarios (e.g., Fig. 1a/2a) was accompanied by true and lure interpretations that were equally plausible, allowing for a more open initial interpretation of the scenario. Interpretations in the IIT were similar in content to those in the emotional BADE task and included 1 Absurd, 2 Lures, and 1 True interpretation for each scenario. To lessen the burden of reading verbal statements, the interpretations in the IIT were shortened to 7 (±2) words, and modified to better describe the scenarios as pictured in the IIT. For example, the IIT scenario depicted in Fig. 1a, b, and c is followed by an absurd interpretation (“People discuss the smell of the field”), two lure interpretations (“People stop you from starting a fight” and “People are making fun of you”), and a true interpretation (“People celebrate
what a great player you are”). Note that the IIT includes two Lure explanations and one True explanation for consistency with BADE tasks (which traditionally have two Lure statements as a remnant of initial attempts to examine the impact of emotional content by comparing emotional vs. neutral Lures). Participants were instructed to imagine explanations and one True explanation for consistency with BADE tasks. Participants were instructed to imagine what a great player you are”). Note that the IIT includes two Lure explanations and one True explanation for consistency with BADE tasks (which traditionally have two Lure statements as a remnant of initial attempts to examine the impact of emotional content by comparing emotional vs. neutral Lures). Participants were instructed to imagine each scenario as if they were an observer in the situation and could see it through their own eyes, given that cognitive biases related to affective symptoms may be more apparent when stimuli are self-relevant (LeMoult et al., 2017).

2.5. Symptom measures

2.5.1. Beck depression inventory-II (BDI-II)

The BDI-II (Beck, Steer, Ball, & Ranieri, 1996) is a well-established 21-item self-report measure of depressive symptom severity over the past two weeks. Participants rate the degree to which they have experienced each symptom on a four-point scale (from 0 to 3). The BDI-II has overall good reliability and validity (Richter, Werner, Heerlein, Kraus, & Sauer, 1998). The internal consistency of the BDI-II in this study was $\omega_{\text{total}} = 0.96$. This measure was included because previous research using the emotional BADE tasks suggests that depressive symptom severity is related to inflexibility of interpretations above and beyond interpretation bias (Everaert et al., 2018; Everaert et al., 2020).

2.5.2. Severity measure for social anxiety disorder – adult (SM-SAD-adult)

The SM-SAD-Adult is a 10-item questionnaire that assesses social anxiety symptomology in adults (Knappe et al., 2013). Participants were asked to report how frequently they have thought, felt, or behaved in particular ways in social situations on a 0–4 scale ranging from “Never” to “All of the time”. This scale has been found to demonstrate good internal reliability (Knappe et al., 2014) and the internal consistency was excellent in this study, $\omega_{\text{total}} = 0.92$. This measure was included because previous research using the emotional BADE tasks suggests that the degree of fear experienced in social situations is related to inflexibility of interpretations above and beyond interpretation bias (Everaert et al., 2018; Everaert et al., 2020).

2.6. Revised Green et al. paranoid thoughts scale (R-GPTS)

The R-GPTS (Freeman et al., 2021) is a 18-item self-report measure of dispositional persecutory ideation (example item: “People have intended me harm”) and self-referential thoughts (example item: “I often heard people referring to me”). Participants rate on a five-point scale (0 = “Not at all,” 4 = “Totally”) how well each item corresponds to their thoughts and feelings about others over the past month. The GPTS demonstrated good psychometric properties, including test-retest reliability and convergent validity (Freeman et al., 2021). Previous research shows that the R-GPTS is sensitive to clinically significant changes in paranoia, discriminates well between individuals with a broad range of latent trait paranoia (Freeman et al., 2021), and has excellent internal consistency in community samples ($\omega_{\text{total}} = 0.96$ in the present study). This measure was included because previous research using the emotional BADE tasks suggests that persecutory ideation is related to inflexibility of interpretations above and beyond interpretation bias (Bronstein et al., 2019).

2.7. Procedure

All participants gave informed consent in accordance with the Institutional Review Board. Participants completed a survey which began with demographic questions followed by the emotional BADE task and the IIT, which were presented in randomized order. Participants then completed the BDI-II, SM-SAD and R-GPTS, which were also presented in a randomized order. Upon completion of the survey, participants were debriefed and received remuneration (6 USD).

2.8. Data reduction and analysis

Emotional BADE task metrics were calculated based on the scoring method employed in prior work (Everaert et al., 2021). Negative and positive interpretation flexibility indices were calculated respectively to represent the ability to revise negative and positive existing beliefs in light of disconfirming evidence. Positive and negative interpretation bias are scored by considering the general tendency to endorse positive or negative interpretations across all stages of the task.

Given that the IIT focuses on resolving ambiguity in gradual steps rather than disconfirming an existing belief, a different computational approach was employed to capture (a) initial interpretation bias and (b) within-person revision of biased interpretations. To this end, an interpretation bias score was calculated at each of the three stages for the 24 IIT scenarios, by taking the true statement plausibility rating divided by the average of the endorsement ratings for the two lures, and multiplying by $-1$ (i.e., additive inverse), so that higher scores represent greater plausibility ratings for lures relative to the true statement. Thus,
the initial interpretation bias was represented by the bias score at stage 1 (first picture of each scenario), the second by the bias score at stage 2 (second picture of each scenario), and the third by the bias score at stage 3 (third picture of each scenario), averaged across scenarios (separately for those with positive and negative outcomes). We then computed an interpretation flexibility index that represents the degree of flexibility in revising prior interpretations based on novel information provided by the IIT scenarios (i.e., stage to stage variability). This was done by taking the average of the differences between the bias scores at stage 3 compared to stage 2 and between stage 2 compared to stage 1. These two differences were squared prior to averaging to capture absolute change (i.e., in either direction), then the square root of the resulting average was computed to bring the index back into the range of the original scaling, as shown below:

\[ \text{Interpretation flexibility index} = \sqrt{\frac{(\text{bias score stage 3} - \text{bias score stage 2})^2 + (\text{bias score stage 2} - \text{bias score stage 1})^2}{2}} \]

This formula, also known as the Root Mean Square of Successive Differences (RMSSD), is the same as used in measuring heartbeat variability and other phenomena that vary across a time series. In this context, the RMSSD is designed to capture stage-to-stage fluctuations in interpretation bias as a proxy for positive or negative interpretation inflexibility (as derived from scenarios with positive or negative emotional valences). Higher RMSSD values represent high stage-to-stage variability and suggest flexibility in revising prior interpretations. By contrast, low RMSSD values represent lower stage-to-stage variability and less flexibility in revising prior interpretations.

The analytic plan aimed to address three objectives: (a) determine the internal consistency of the IIT, (b) determine the degree of convergence between the emotional BADE and IIT indicators, and (c) determine the utility of the IIT indicators to explain variation in symptoms of depression, social anxiety and paranoia. First, internal consistency for all indices in both the emotional BADE task and IIT was calculated using McDonald's Omega (Hayes & Couts, 2020). Second, to confirm that the IIT is measuring similar constructs of interpretation flexibility as the emotional BADE, the study examined convergent validity of the IIT and emotional BADE task indices using Pearson correlations. Last, the present study aims to replicate previously observed relations between emotional BADE indices and different psychiatric symptom dimensions, with a particular focus on converging findings between the emotional BADE and newly developed IIT. Multiple regression models were built to test whether interpretation bias and inflexibility uniquely explained variation in depression, social anxiety, and delusion-like ideation. For each task and scenario type, regression models were tested separately for depression, social anxiety, and paranoia as dependent variables. For the emotional BADE task, in each model, negative and positive interpretation bias, as well as interpretation flexibility (negative or positive, depending on the scenario type) were simultaneously entered into the regression equation. Each model for the IIT indices included interpretation flexibility index and initial interpretation bias, in positive and negative outcome scenarios, respectively, as the independent variables.

### 3. Results

#### 3.1. Sample characteristics

Participants’ BDI-II scores (M = 14.80, SD = 14.45) represented variation across the spectrum of depressive symptom severity: 149 respondents reported minimal (range: 0–13), 31 reported mild (range: 14–19), 42 reported moderate (range: 20–28), and 51 reported severe (range: 29–58) depressive symptoms. Significant variation was also found in participants’ scores on measures of social anxiety (M = 7.88, SD = 8.98, range: 0–39), and delusion-like ideation (M = 9.28, SD = 13.47, range: 0–72). The observed symptom scores for the BDI-II, SM-SAD, and the R-GPTS enabled this study to examine the relations between interpretation bias/inflexibility and psychopathology along broad continua of symptom severity.

#### 3.2. Scenario structure check

Participants exhibited a general trend of revising their interpretations in both the IIT (Fig. 2a) and Emotional BADE task (Fig. 2b), with the mean likelihood ratings for lures decreasing (i.e., a decreased tendency to endorse lures as the scenarios were increasingly disambiguous), and the ratings for true interpretations increasing across three stages of the task. By the end of the scenarios, True explanations were rated as being significantly more plausible than Lure explanations (t = 47.56, p < .001 for IIT; t = 56.48, p < .001 for the Emotional BADE task). This pattern of changes in likelihood ratings is evident in both negative and positive outcome scenarios of the IIT, highlighting the validity of the unblurring task procedure in prompting revision of initial interpretations. Consistent with the structure of the Emotional BADE task, absurd interpretations on the IIT remained implausible throughout the scenarios (mean rating: 1.48, SD = 0.56), and were perceived as being less plausible than either the Lure (mean rating: 2.90, SD = 0.58) or True (mean rating 4.14, SD = 0.61) explanations (ps < .001).

#### 3.3. Convergence between IIT and emotional BADE task indices

Both the IIT and emotional BADE task had an excellent internal consistency. The internal consistency coefficient (McDonald’s Omega total) was \( \omega = 0.90 \) for absurd scenarios, \( \omega = 0.87 \) for lure scenarios and \( \omega = 0.83 \) for true scenarios of the IIT. The emotional BADE task has \( \omega = 0.95 \) for absurd scenarios, \( \omega = 0.95 \) for lure scenarios and \( \omega = 0.91 \) for true scenarios.

Indices for both interpretation bias and inflexibility were significantly correlated across the emotional BADE and IIT (Table 2). As hypothesized, metrics of negative interpretation bias (\( r = 0.47, df = 245, p < .001 \)) and positive interpretation bias (\( r = 0.34, df = 245, p < .001 \)) derived from the IIT and Emotional BADE Task were positively correlated. Similarly, there was a moderate correlation between metrics of interpretation inflexibility derived from the two tasks (for positive interpretation inflexibility: \( r = -0.49, df = 245, p < .001 \), for negative interpretation inflexibility: \( r = -0.38, df = 245, p < .001 \)). These correlations are indicative of acceptable convergent validity (Fernandez-Marcos, de la Fuente, & Santacreu, 2018), especially given that the emotional BADE task focuses on revision of interpretations in response to disconfirmatory evidence, whereas the IIT focuses on resolution of ambiguity.

#### 3.4. Explaining variation in psychopathology symptoms

Assumptions of homoscedasticity and normality of residuals were met for all analyses. Collinearity statistics were within acceptable limits (VIFs < 1.21, Tolerance’s > 0.83). Regression models for the emotional
a. IIT

Results from the disconfirming-the-positive portion of the emotional BADE task and the negative outcome scenarios of the IIT are presented below, followed by those from the disconfirming-the-negative portion of the emotional BADE task and the positive outcome scenarios of the IIT.

3.4.1. Negative outcome scenarios

In the emotional BADE task, disconfirming-the-positive scenarios initially encourage a positive interpretation, but as events in the scenario unfold, a more negative interpretation becomes warranted. Similarly, the negative outcome scenarios of the IIT reveal information about initially ambiguous scenarios by unblurring the pictorial stimuli, leading to a more definitive interpretation.

b. Emotional BADE

Table 2
Correlation Matrix between the IIT and Emotional BADE Indices.

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. PIB (IIT)</td>
<td>–0.39**</td>
<td>–0.09</td>
<td>0.05</td>
<td>0.34**</td>
<td>–0.31**</td>
<td>0.22*</td>
<td>0.10</td>
</tr>
<tr>
<td>2. NIB (IIT)</td>
<td>–</td>
<td>0.20**</td>
<td>–0.03</td>
<td>–0.33**</td>
<td>0.47**</td>
<td>–0.44**</td>
<td>–0.17**</td>
</tr>
<tr>
<td>3. Positive IFI (IIT)</td>
<td>–</td>
<td>0.36**</td>
<td>–0.15*</td>
<td>0.15*</td>
<td>–0.50**</td>
<td>–0.47**</td>
<td></td>
</tr>
<tr>
<td>4. Negative IFI (IIT)</td>
<td>–</td>
<td>–0.05</td>
<td>–0.01</td>
<td>–0.32**</td>
<td>–0.38**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. PIB (BADE)</td>
<td>–</td>
<td></td>
<td></td>
<td>0.21**</td>
<td>0.53**</td>
<td></td>
<td>0.17**</td>
</tr>
<tr>
<td>6. NIB (BADE)</td>
<td>–</td>
<td></td>
<td></td>
<td>–0.19**</td>
<td>–0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. PII (BADE)</td>
<td>–</td>
<td></td>
<td></td>
<td>–0.81**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. NII (BADE)</td>
<td>–</td>
<td></td>
<td></td>
<td>–0.81**</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. PIB = positive interpretation bias, NIB = negative interpretation bias, IFI = interpretation flexibility index, PII = positive interpretation inflexibility, NII = negative interpretation inflexibility.

*, p < .05.
**, p < .01.

BADE task and IIT are presented in Tables 3 and 4, respectively. Results from the disconfirming-the-positive portion of the emotional BADE task and the negative outcome scenarios of the IIT are presented below, followed by those from the disconfirming-the-negative portion of the emotional BADE task and the positive outcome scenarios of the IIT. We hypothesized parallel patterns of association between BADE and IIT indices with the psychopathology measures within each type of outcome scenario. Results from scenarios that resolved to a negative outcome (disconfirming-the-positive scenarios from the Emotional BADE task, negative outcome scenarios from the IIT) are presented first below, followed by results from scenarios featuring a positive outcome (disconfirming-the-negative scenarios from the Emotional BADE Task, positive outcome scenarios from the IIT).
3.4.1. Emotional BADE. A regression analysis indicated that the three emotional BADE components explained a significant amount of variance in depressive symptoms (F(3, 269) = 12.67, p < .001, R² = 0.12), social anxiety (F(3, 269) = 14.66, p < .001, R² = 0.14), and delusion-like beliefs (F(3,269) = 14.66, p < .001, R² = 0.14). Positive interpretation bias was associated with lower depression severity (β = −0.29, t(272) = −4.05, p < .001) and less severe social anxiety symptoms (β = −0.33, t(272) = −4.71, p < .001), but not delusion-like belief (β = −0.13, t(272) = −1.87, p = .063). Negative interpretation bias was linked to more severe symptoms of depression (β = 0.36, t(272) = 5.80, p < .001) and social anxiety (β = 0.48, t(272) = 8.07, p < .001) and more delusion-like ideation (β = 0.28, t(272) = 4.56, p < .001). Additionally, a higher level of depressive symptoms (β = 0.24, t(272) = 3.31, p = .001), social anxiety (β = 0.31, t(272) = 4.49, p < .001) and delusion-like ideation (β = 0.40, t(272) = 5.54, p < .001) was related to greater positive interpretation inflexibility.

### Table 3
Regression Models with the Emotional BADE Task Indices Predicting Psychiatric Symptoms.

<table>
<thead>
<tr>
<th>Scenario Component</th>
<th>b</th>
<th>SE</th>
<th>β</th>
<th>t</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Depression</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>27.29</td>
<td>2.85</td>
<td>9.57</td>
<td>3.32</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Negative outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>30.78</td>
<td>2.82</td>
<td>10.94</td>
<td>3.32</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>8.17</td>
<td>3.08</td>
<td>2.65</td>
<td>4.49</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Negative outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>9.00</td>
<td>2.72</td>
<td>3.17</td>
<td>9.57</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Delusions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>13.26</td>
<td>1.88</td>
<td>7.04</td>
<td>3.32</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Negative outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>11.27</td>
<td>2.35</td>
<td>5.04</td>
<td>2.50</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Note. IFI = interpretation flexibility index, NIB = negative interpretation bias, NIB = negative interpretation bias.

### Table 4
Regression Models with the IIT Indices Predicting Psychiatric Symptoms.

<table>
<thead>
<tr>
<th>Scenario Component</th>
<th>b</th>
<th>SE</th>
<th>β</th>
<th>t</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Depression</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>23.26</td>
<td>3.03</td>
<td>7.81</td>
<td>3.03</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Negative outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>4.97</td>
<td>2.52</td>
<td>0.12</td>
<td>4.56</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>13.26</td>
<td>1.88</td>
<td>7.04</td>
<td>3.03</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Negative outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>11.27</td>
<td>2.35</td>
<td>5.04</td>
<td>2.50</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Delusions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>30.78</td>
<td>2.82</td>
<td>10.94</td>
<td>3.32</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Negative outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>27.26</td>
<td>2.85</td>
<td>9.57</td>
<td>3.32</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Note. IFI = interpretation flexibility index, PIB = positive interpretation bias, NIB = negative interpretation bias.

the participants to make negative interpretations about the scenarios by the last picture.

3.4.1.2. IIT. The results of the regression showed that the two IIT scores also explained a significant amount of the variance in depressive symptoms (F(2, 270) = 3.91, p = .021, R² = 0.03), social anxiety (R² = 4.52, p = .012, R² = 0.03), and delusion-like beliefs (F(2, 270) = 5.775, p = .003, R² = 0.04). In particular, more positive interpretation bias was linked to less severe depressive symptoms (β = −0.17, t(272) = −2.80, p = .063), less severe social anxiety (β = −0.18, t(272) = −3.00, p = .003) and less delusion-like ideation (β = −0.13, t(272) = −2.15, p = .032). While a higher level of delusion-like ideation was related to greater interpretation inflexibility (β = −0.17, t(272) = −2.81, p = .005), no significant relationship was found between interpretation inflexibility and depression severity (β = −0.01, t(272) = −0.12, p = .907) or social anxiety severity (β = −0.01, t(272) = −0.13, p = .895).

In sum, in negative outcome scenarios, depression and social anxiety were related to both interpretation bias and inflexibility in the emotional BADE task. In the IIT, however, depression and social anxiety were only associated with positive interpretation bias, not inflexibility. Delusion-like ideation, however, was related to interpretation inflexibility in both tasks. In a follow-up analysis, two sets of models were built: one with the flexibility indices derived from the IIT and Emotional BADE task as predictors for symptoms; the other with a composite variable (i.e., the absolute sum of flexibility indices from the IIT and Emotional BADE) as the predictor, which forces a constraint on the coefficient to be equal. In comparing the model fit with an ANOVA test, no significant differences were found between the two models in predicting symptoms of depression (F = 0.05, p = .83), social anxiety (F = 0.38, p = .54) and delusions (F = 3.18, p = .008). In other words, the amount of variance in symptoms as explained by the flexibility indices in the IIT and Emotional BADE task are not significantly different.

3.4.2. Positive outcome scenarios

In the emotional BADE task, confirming-the-negative outcome scenarios initially encourage a negative interpretation, but as events in the scenario unfold, a more positive interpretation becomes warranted. Similarly, the positive outcome scenarios of the IIT reveal information about initially ambiguous scenarios by unblurring the pictorial stimuli, leading the participants to arrive at positive interpretations about the scenarios by the last picture.

3.4.2.1. Emotional BADE. The three emotional BADE components explained a significant amount of variance in depressive symptoms (F(3, 269) = 12.48, p < .001, R² = 0.12), social anxiety (F(3, 269) = 23.26, p < .001, R² = 0.21), and delusion-like ideation (F(3, 269) = 12.91, p < .001, R² = 0.13). While more positive interpretation bias was linked to less severe depressive symptoms (β = −0.19, t(272) = −3.17, p = .002) and social anxiety (β = −0.19, t(272) = −3.38, p = .001), there was no significant relationship between positive interpretation bias and delusion-like ideation (β = 0.04, t(272) = 0.71, p = .477). More negative interpretation bias was associated with greater depression severity (β = −0.31, t(272) = 5.32, p < .001), social anxiety severity (β = −0.42, t(272) = 7.45, p < .001), and more delusion-like ideation (β = 0.20, t(272) = 3.17, p = .001).
3.37, \( p = .001 \). Greater negative interpretation inflexibility is associated with more severe depressive symptoms (\( \beta = 0.19, t(272) = 3.23, p = .001 \)), social anxiety (\( \beta = 0.25, t(272) = 4.53, p < .001 \)), and higher level of delusion-like ideation (\( \beta = 0.30, t(272) = 5.07, p < .001 \)).

3.4.2.2. IIT. Regression analysis indicated that the IIT indices explained a significant proportion of the variance in depressive symptoms (F(2, 270) = 5.47, \( p = .005, R^2 = .04 \)), social anxiety (F(2, 270) = 5.10, \( p = .007, R^2 = 0.04 \)), and delusion-like ideation (F(2, 270) = 6.247, \( p = .002, R^2 = 0.044 \)). In particular, a greater level of negative interpretation bias (\( \beta = 0.12, t(272) = 1.97, p = .050 \)) and greater interpretation inflexibility (\( \beta = -0.16, t(272) = -2.61, p = .010 \)) were associated with more depression severity. Similarly, more negative interpretation bias (\( \beta = 0.12, t(272) = 1.99, p = .047 \)) and greater interpretation inflexibility (\( \beta = -0.15, t(272) = -2.45, p = .015 \)) were linked with a greater level of social anxiety symptoms. While higher level of delusion-like ideation was related to greater interpretation inflexibility (\( \beta = -0.21, t(272) = -3.51, p = .001 \)), the association between negative interpretation bias and delusion-like ideation was not significant in the IIT (\( \beta = -0.03, t(272) = -0.54, p = .590 \)).

Overall, in positive outcome scenarios, depression and social anxiety were related to negative interpretation bias and interpretation inflexibility in both the BADE and IIT tasks. Delusion-like ideation was consistently associated with interpretation inflexibility across both tasks, but was not consistently linked with interpretation bias. In a follow-up analysis comparing the amount of variance explained by flexibility indices in the IIT vs. Emotional BADE task, no significant differences were found between the IIT and BADE models in predicting symptoms of depression (\( F = 0.19, p = .67 \)), social anxiety (\( F = 0.26, p = .61 \)) and delusions (\( F = 1.44, p = .23 \)).

4. Discussion

The present study examined the psychometric properties of a novel task assessing interpretation bias and flexibility when processing emotional stimuli, and investigated the relations between task performance and psychopathology. Metrics derived from the IIT exhibited good-to-excellent internal consistency, implying that the majority of inter-individual variance in these metrics is due to variation in interpretive processes. Moreover, these metrics exhibited convergent validity with existing measures of interpretation bias and inflexibility. Specifically, metrics of interpretation bias and inflexibility derived from the IIT were significantly correlated with those derived from the emotional BADE task, and across tasks a similar pattern of associations emerged between these metrics and symptoms of psychopathologies (depression, social anxiety, paranoia) previously associated with interpretation inflexibility (Everaert et al., 2018). The convergence between tasks is especially notable given that the lack of convergent validity in commonly used measures of interpretation bias has likely contributed to the inconsistent findings across existing studies of interpretive processes (O'Connor et al., 2021).

As one of the first studies to investigate the relation between interpretation inflexibility, affective symptoms, and persecutory ideation in the same sample, the current study showed important similarities and differences across diagnostic dimensions. Both the IIT and Emotional BADE task suggested that depression, social anxiety, and persecutory ideation are all associated with greater interpretation inflexibility. The fact that interpretation inflexibility was consistently associated with symptom profiles regardless of stimulus modality points toward a higher-order cognitive deficit (i.e., one not specific to any perceptual information-processing stream) that confers inflexibility in interpretive processes and may thereby increase risk of experiencing symptoms of multiple psychopathologies. Decreased ability to detect response conflict arising from situational ambiguity, and problems implementing sufficient reactive cognitive control to support interpretation updating are candidate mechanisms that may contribute to this inflexibility. Broadly consistent with this possibility, individuals who exhibit depression and anxiety, as well as those prone to paranoia, exhibit deficits in conflict processing (Holmes & Pizzagalli, 2008; Krug & Carter, 2010; Speechley, Woodward, & Ngan, 2013). Other measures of cognitive ability, such as working memory capacity, may also relate to interpretation inflexibility in emotional scenarios. Future research could examine the role of these deficits as a first step toward determining why people with depression, social anxiety, and paranoia exhibit inflexibility across the IIT and Emotional BADE tasks.

Interestingly, interpretation inflexibility in depression appears to be stronger in the context of revising negative interpretations—a pattern that is shown in both previous emotional BADE studies (Everaert et al., 2018) and the current IIT findings, and one which is consistent with other depression research (e.g., studies showing an association between depression and biased reappraisal of disconfirmatory information that maintains negative expectations and interpretations; Kube et al., 2019). In contrast, such inflexibility in paranoid individuals appears to exist regardless of the emotional valence of initial beliefs (Brotho et al., 2019). Thus, while interpretation inflexibility is related to multiple forms of psychopathology, its dependence on the emotional content of the information being interpreted, and the mechanisms underlying the inflexibility, may differ somewhat across symptom domains. Given the relevance of emotional valence to inflexibility in depression and social anxiety, future research should examine whether other emotion-related properties of the information being interpreted (location in valence-arousal space, for example) impacts inflexibility.

Future research could also take advantage of the unique properties of the IIT, resulting from its reliance on picture-based scenarios, to elucidate mechanisms that influence various psychopathologies. For example, only the IIT requires individuals to employ facial emotion recognition to resolve ambiguity, and the IIT (vs. the Emotional BADE Task) may more heavily depend on theory of mind. The IIT may therefore be more suitable for future research investigating whether these deficits can explain interpretation bias and inflexibility in individuals with depression, social anxiety, or paranoia. The IIT also eliminates the confound of varying reading comprehension levels across different developmental stages, making it preferable for developmental studies examining whether inflexible or biased interpretations precede and predict psychopathologies. For example, the IIT could be used to investigate the temporal priority of inflexibility over suicide ideation, which was linked to interpretation inflexibility in a recent, longitudinal study of adults (Everaert et al., 2021). The IIT is also preferable for research using electrophysiological or functional magnetic resonance imaging to uncover biobehavioral markers of interpretation bias and inflexibility, as picture-based stimuli may afford more precise stimulus timing. Finally, the IIT may be more useful for studies examining different cultural contexts and diverse social backgrounds, especially for non-English-speaking groups who are likely to be traditionally underrepresented in study samples (Oh et al., 2015), because it reduces the challenges posed by language barriers. To lay foundations for such applications, additional research is needed to examine the impacts of cultural factors on the interpretation of the photos, with particular attention to whether culture alters their perceived levels of ambiguity.

The study is not without limitations. Participants recruited from an online crowd-sourcing platform captured a sample with elevated symptoms based on self-report. While this recruitment method has been shown to produce valid data comparable to those collected in other settings (e.g., in-person lab environment), a more stringent recruitment process (potentially including clinical interviews) is needed to provide information on the role of interpretation inflexibility in clinical populations. Moreover, while the current study presented a broad examination of interpretation inflexibility in relation to affective symptoms and persecutory ideation, the study did not account for the shared variance between depression and social anxiety. Future research should
further examine the specificity of depression versus social anxiety by modeling the relative impact of these conditions using statistical techniques (e.g., commonality analysis). To better capture the shared vs. distinctive cognitive mechanisms among symptom dimensions, future studies could design a more fine-grained recruitment process that groups participants based on the severity of different symptoms. Considering the prevalent comorbidity across symptoms, more work is needed to better understand how symptoms may interact with one another and the implications of interpretation inflexibility on such comorbidity. Additionally, the cross-sectional design precludes conclusions about causal relations. Longitudinal studies are needed to investigate the cross-lagged relations between interpretation bias or interpretation inflexibility and symptoms of various forms of psychopathology. Given the emphasis of the IIT on processing new and ambiguous information, a longitudinal design also allows participants to become less familiar with the scenarios (that were disambiguated during their initial exposure to the task) before completing the task again, so as to help establish test-retest reliability.

Finally, while not a limitation per se, readers should keep in mind that the types of interpretation inflexibility assessed by the emotional BADE task and the IIT differ: the IIT examines rigidity in maintaining a particular level of interpretation bias as scenarios are disambiguated, whereas the emotional BADE task examines rigidity in maintaining an initial interpretation of an ambiguous situation in the face of evidence against it. While this nuance demands attention when synthesizing findings across tasks, it suggests that comparing performance on the IIT and emotional BADE tasks in people with various kinds of psychopathology can provide tantalizing hints about the particular types of inflexibility that might most strongly contribute to these individuals' experiences of distress and disability. In line with investigating specific types of inflexibility, future studies may also assess the discriminant validity of the IIT (and emotional BADE task) – for example, the inflexibility indices from the IIT and emotional BADE may be more highly correlated given that they both capture the interpretation inflexibility, whereas tasks or questionnaires measuring other flexibility constructs (e.g., psychological flexibility or task-switching ability) may be less strongly associated with the IIT indices.

In conclusion, the IIT is a reliable paradigm to capture interpretation biases and flexibility when processing emotional stimuli. The novel use of pictorial stimuli boosts the ecological validity and task applicability across a wide range of research contexts. Future research with the IIT will further our understanding of interpretation inflexibility as a trans-diagnostic mechanism underlying symptoms across the severity spectrum.

CRediT authorship contribution statement

Wisteria Deng: Conceptualization, Data curation, Formal analysis, Writing- Original draft preparation.
Jonas Everaert: Conceptualization, Methodology, Writing- review & editing.
Mackenzie Creighton: Data curation, Project administration. Michael Bronstein: Methodology, Writing- review & editing.
Tyrone Cannon: Conceptualization, Methodology, Writing- review & editing.
Jutta Joormann: Conceptualization, Methodology, Writing- Reviewing and Editing.

Declaration of competing interest

The authors declare no conflict of interest and the authors received no funding from an external source.

References

reliability and sensitivity to change of the dimensional anxiety scales for DSM-5. CNS Spectrums, 19(3), 256–267. https://doi.org/10.1017/S1092852919300710
