**Feng, YC, Krahé, C, Sumich, A, Meeten, F, Lau, JYF and Hirsch, CR**

**Using event-related potential and behavioural evidence to understand interpretation bias in relation to worry**


# Using event-related potential and behavioural evidence to understand interpretation bias in relation to worry

Ya-Chun Feng, Charlotte Krahé, Alexander Sumich, Frances Meeten, Jennifer Y.F. Lau, Colette R. Hirsch

PII: S0301-0511(18)30665-3

DOI: https://doi.org/10.1016/j.biopsycho.2019.107746

Article Number: 107746

Reference: BIOPSY 107746

To appear in:

Received Date: 7 September 2018

Revised Date: 3 August 2019

Accepted Date: 26 August 2019

Please cite this article as: Feng Y-Chun, Krahé C, Sumich A, Meeten F, Lau JYF, Hirsch CR, Using event-related potential and behavioural evidence to understand interpretation bias in relation to worry, *Biological Psychology* (2019), doi: https://doi.org/10.1016/j.biopsycho.2019.107746

# Using event-related potential and behavioural evidence to understand interpretation bias in relation to worry

Running head: INTERPRETATION BIAS IN RELATION TO WORRY

Ya-Chun Feng[a], Charlotte Krahé[a1], Alexander Sumich[b, c], Frances Meeten[a2], Jennifer Y. F. Lau[a], & Colette R. Hirsch[a]

[a] Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

[b] Division of Psychology, School of Social Sciences, Nottingham Trent University, Nottingham, UK

[c] Department of Psychology, Auckland University of Technology, Auckland, NZ

Correspondence: Dr Colette Hirsch, Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF, Email: colette.hirsch@kcl.ac.uk

[1]Now at Department of Psychological Sciences, University of Liverpool, Liverpool, UK.

[2]Now at School of Psychology, University of Sussex, Falmer, Brighton, UK.

Hightlights of manuscript: “Using event-related potential and behavioural evidence to understand interpretation bias in relation to worry”

- We found high and low worriers have different interpretation tendencies
- High worriers lack the benign interpretation bias evident in low worriers
- As evidenced by event-related potential reflecting initial online interpretations
- Also, in behavioral indices that measure the interpretation after reflection
- Event-related potential can be a sensitive method to assess interpretation bias

Abstract

The tendency of interpreting ambiguous information in a consistent (e.g. negative) manner (interpretation bias) may maintain worry. This study explored whether high and low worriers generate different interpretations and examined at which stages of information processing these interpretations can occur. Participants completed interpretation assessment tasks yielding behavioural and N400 event-related potential indices, which index whether a given interpretation was generated. High worriers lacked the benign interpretation bias found in low worriers. This was evident for early "online" interpretations (reflected in reaction times to relatedness judgments and lexical decisions, as well as at a neurophysiological level, N400, for lexical decisions only), to later "offline" interpretations (observed at a behavioural level on the scenario task and recognition task) when participants had time for reflection. Results suggest that a benign interpretation bias may be a protective factor for low worriers, and that these interpretations remain active across online and offline stages of processing.

**Keywords**: interpretation bias; worry; N400; online interpretation; offline interpretation

# Using event-related potential and behavioural evidence to understand interpretation bias in relation to worry

Worry is a stream of negative thoughts about potential future outcomes and is experienced by many individuals in the general population (Barlow, 2002). Although many people (28%) experience excessive and uncontrollable worry, only a minority of people (6%) meet the criteria of generalized anxiety disorder (GAD; Ruscio, 2002). Given that excessive worry is such a common experience, it is important to understand how it is maintained in order to develop effective management strategies. This paper aims to understand whether interpretation biases, which are one of the factors proposed to maintain worry (Hirsch & Mathews, 2012), differ between individuals with high and low worry levels. Furthermore, it aims to understand the time course at which such biases occur using behavioural and neurophysiological measures to explore at which stages of information processing differences in interpretations are evident.

## The role of interpretation in worry

Streams of worry often involve ambiguity or uncertainty. If a negative interpretation is generated, this can lead to further negative worrying thoughts (e.g., I will not have enough money to pay for the bill), and generate more negative interpretations. In contrast, generating benign interpretations (e.g., I will have saved enough money) could terminate bouts of worry. The consistent tendency to interpret information in a certain (e.g., negative) manner is referred to as "interpretation bias". Previous studies have suggested that interpretation bias plays a causal role in maintaining worry, evidenced by reduced worry (indexed by fewer negative intrusions) when individuals were asked to concentrate on their breathing (breathing focus task; Hayes, Hirsch, Krebs, & Mathews, 2010; Hirsch, Hayes, & Mathews, 2009), and lower self-reported levels of worry (Hirsch et al., 2018) after a cognitive bias modification for interpretation (CBM-I) program, which provides repeated practice in generating benign inferences about ambiguous situations. Although interpretation bias seems to play an important

role in the maintenance of worry, previous studies have focused on interpretation bias in relation to GAD, rather than worry *per se* (e.g., Anderson et al., 2012; Eysenck, Mogg, May, Richards, & Mathews, 1991; Mogg, Baldwin, Brodrick, & Bradley, 2004).

To date, only two studies have focused on assessing interpretation bias in worry, both of which involved accessing offline interpretations, that is, interpretations made after opportunity for reflection. One study focused on a non-clinical child cohort of high and low trait worriers (Suarez & Bell-Dolan, 2001). The other study explored the relationship between levels of worry and negative interpretation bias in adults with GAD and healthy controls (Krahé, Whyte, Bridge, Loizou & Hirsch, 2019). Both studies showed that higher levels of worry were associated with a greater negative interpretation bias. However, understanding interpretation bias in adults with high levels of worry remains relatively unexplored since Suarez & Bell-Dolan (2001) used children sample and Krahé et al. (2019) used individuals with GAD as a sample. Although Suarez & Bell-Dolan (2001) compared groups of high and low worriers, they did not distinguish between the presence of negative interpretation bias and an absence of benign interpretation bias, which is important when developing strategies to lessen worry. If individuals have a negative interpretation bias, then reducing this negative bias may benefit individuals; while facilitating benign bias may be more beneficial to individuals when they lack a benign interpretation bias.

Another issue here is that both studies only investigated interpretations after some time of reflection; however, interpretations can be generated more spontaneously and earlier than upon reflection. Understanding when interpretations are generated and whether interpretation biases are consistent at different time courses is essential in constructing the cognitive model of worry and has implications in terms of intervening to facilitate positive interpretation bias. If negative interpretation bias only occurs on reflection, then individuals could actively challenge the interpretations during a period of reflection. However, if negative interpretation bias occurs

spontaneously when first encountering information, then techniques designed to promote a more positive interpretation bias may be most efficient if they specifically target changing interpretations generated in the early stages of processing.

**"Offline" and "online" interpretations**

As mentioned, interpretations can be generated at different stages of information processing: after reflecting on the ambiguous information (offline), or when individuals first encounter ambiguous information (online; Hirsch, Meeten, Krahé, & Reeder, 2016). A limitation in the previous research that investigated interpretation bias is that they have only investigated either offline interpretations (e.g., Amir, Foa, & Coles, 1998; Anderson et al., 2012), or online interpretations (e.g., Hirsch & Mathews, 1997; 2000), rather than exploring interpretations at different information processing stages in the same study. Furthermore, most previous studies have mainly focused on offline interpretation bias in individuals with high levels of anxiety (e.g., Amir et al., 1998; Anderson et al., 2012), who often experience uncontrollable worry about certain topics (e.g., social topics for socially anxious populations, several general topics for GAD) and are accompanied by somatic symptoms, such as restlessness, fatigue, and sleep disturbance (APA, 2013). For example, researchers have asked anxious participants to read ambiguous situations and to rate their concerns about the situations (Anderson et al., 2012), or rank order potential negative, neutral, and positive interpretations in terms of how likely they are to come to mind (Amir et al., 1998; Stopa & Clark, 2000, Voncken, Bögels, & de Vries, 2003). These studies found that anxious groups were more concerned about the ambiguous situations and, compared to control groups, rated negative interpretations as more likely to come to their mind. Other offline paradigms showed similar effects when participants were required to spell words that included homophones, which had both negative and benign meanings (e.g., die, dye; Mathews, Richards, & Eysenck, 1989; Mogg et al., 2004). Anxious participants wrote down more negative spellings of homophones

than controls, indicating again that anxious participants demonstrated a negative offline interpretation bias.

The above offline interpretation bias measures that involve ranking potential interpretations provide the relative positioning (e.g., likelihood ranking), or rating on one dimension (e.g., level of concern), do not enable us to assess the extent to which individuals make negative and/or benign interpretations, respectively. Furthermore, such measures do not enable us to understand the relative bias within a group, that is, whether a group generates one type of interpretation (e.g., negative, benign) more than another. For example, if anxious individuals rated situations as more concerning than controls, this could imply that anxious individuals have a greater negative interpretation bias, or that they generate fewer benign interpretations than controls. The relative bias (e.g., whether they make more negative than benign interpretations) cannot be determined by this rating.

In contrast, the "recognition task" is an offline task that presents independent ratings for negative and benign interpretations, which enable us to explore the relative bias in a group (Eysenck et al, 1991). In the recognition task, participants read ambiguous scenarios and are then instructed to rate how similar the benign or negative statements are to the original scenarios. In doing so, the similarity ratings for benign and negative statements provide negative and benign interpretation indices. Eysenck et al. (1991) found that a non-anxious group endorsed more benign interpretations than negative interpretations (benign bias), while a GAD group did not show any difference between the two types of interpretations (no bias). Thus, anxious individuals lacked the normal benign interpretation bias found in the general population when there is an opportunity to reflect on ambiguous information, which is different from previous studies that did not measure relative bias differences (e.g., Anderson et al., 2012).

Online interpretation measures typically provide both negative and benign interpretation

indices and, as a result, a relative bias can be examined. Online tasks involve responding to stimuli that resolve ambiguity in a negative or benign way as quickly as possible; faster reaction times indicate that the individual has generated a matching interpretation. A further benefit of online interpretation measures is that they are less subject to experimental demand, because the goals are more opaque (Hirsch et al., 2016). Two commonly used online interpretation bias measures are the lexical decision task (LDT) and the word sentence association task (WSAT). The LDT requires participants to read ambiguous scenarios (e.g., Your new job has changed your life for the___), and then to determine as quickly as possible whether or not the final word of the ambiguous scenarios is a real word (e.g., better/worse; Hirsch & Mathews, 1997; 2000). The target word resolves the ambiguity in a negative or benign manner. If the word matches the interpretation that the individual has generated, it is easier for them to recognize it as a real word, so they will respond faster than if the word does not match their interpretation. The reaction time for different trial types (benign or negative) can be calculated independently and the within-group negative interpretation bias in these tasks are reflected in faster reaction time to negative than to benign trials. In the Hirsch and Mathews' studies (1997; 2000), non-anxious participants responded faster to benign than negative trials, but socially anxious participants responded equally quickly to benign and negative trials. This indicates that the non-anxious group had a benign interpretation bias, while socially anxious individuals lacked this benign bias and were un-biased in their interpretations. To date, no study has compared online relative interpretation bias between high and low worriers. Therefore, the current study used online tasks to explore whether individuals with high levels of worry also lack of the benign bias that low worriers have, as in the two Hirsch and Mathews' studies (1997; 2000).

Another commonly used online task is the WSAT (Amir, Prouvost & Kuckertz, 2012; Beard & Amir, 2009), which can also be used to investigate a relative bias. It requires participants to determine whether or not words (e.g., warning, bottle) are related to the

following sentences (e.g., The alarm goes off) as quickly as possible. The words in this task also resolve the sentences in a benign or negative manner (e.g., clock, warning). Two studies used this measure (Amir et al., 2012; Beard & Amir, 2009) and found that non-anxious participants endorsed words matching benign interpretations more quickly than socially anxious participants, showing that non-anxious participants had a greater benign interpretation bias. Another study (Oglesby, Raines, Short, Capron, & Schmidt, 2016) used this measure to explore interpretation bias in relation to intolerance of uncertainty, which is a characteristic resulting from negative beliefs about uncertainty and is highly related to anxiety. They found that higher levels of intolerance of uncertainty were associated with a greater negative interpretation bias. The above WSAT studies have only compared benign and negative interpretation biases independently across different levels of anxiety or intolerance of uncertainty, but did not explore the relative bias of a specific group. Therefore, interpretation bias within a group is still unclear and is explored more in the current study.

**Measuring online interpretations by electroencephalography (EEG)**

Although using online interpretation measures can explore interpretations made at the point at which individuals first encounter ambiguity, reaction times may not be the most immediate and direct index to measure online interpretations since they still involve multiple cognitive and behavioural processes (e.g., selecting a response, executing a button press etc.). In the LDT, a behavioural response is required after resolving ambiguity and making a lexical decision, in the order of seconds (Hirsh & Mathews, 1997; 2000). In contrast, neurophysiological measures, such as electroencephalography (EEG), provide information with millisecond precision of the cognitive processes (i.e., at the moment when ambiguity is resolved) that occur prior to a behavioural response and are thus a more sensitive index than reaction time.

EEG activity that is "time-locked" to the presentation of a stimulus (e.g., target) can be measured by event-related potentials (ERPs), i.e., evoked brain responses, which are represented by a series on positive (P) and negative (N) components. In the online tasks mentioned above, individuals' interpretations were measured by the reaction time that indicates whether the target word was in line with individuals' expectations when resolving ambiguity (i.e., faster reaction time when in line with interpretation). In regards to endorsing – and violating – expectations, a negative component occurring approximately 400ms post-stimulus, namely the N400, has previously been studied. The N400 reflects the ease with which semantic information is assessed and integrated into a given context (Kutas & Federmeier, 2000; Swaab, Ledoux, Camblin, & Boudewyn, 2012). The ease of integration is based on individuals' semantic memory, which is constructed by knowledge through experiences. If individuals have built up a negative knowledge set around one context, they will expect the outcomes for that context (or other similar contexts) to be negative, so it will be easier for them to integrate negative information into that context. Larger negative N400 amplitudes are elicited when facing information that is more difficult to integrate into the preceding context (i.e., when it is not expected), which would occur when a different interpretation has been generated. Therefore, N400 can be used to index interpretations in online tasks by signalling whether the target word resolves ambiguity in keeping with the individuals' expectation (i.e., it matches their interpretation) or not. If the target word does not match the participant's interpretation, it will violate expectation, leading to a greater N400 compared to a target word in keeping with expectation. Moreno and colleagues (Moreno & Rivera, 2014; Moreno &Vázquez, 2011) presented sentences with emotional outcome to samples from the general population. The final words of each sentences decided whether the sentence were negative, positive (both studies), or neutral (only Moreno & Rivera, 2014). They found the N400 effects on high vs. low expectation final words were different between emotional sentences, indicating N400 can be modulated by different emotions. Exploring how ambiguity is resolved (e.g. in a benign or

negative way) in high and low worriers in the processing stage presented by N400 would be helpful in terms of understanding what the "online" interpretive tendency that may contribute to worry is, and whether this tendency changes in the later "offline" information processing stages or remains the same from the "online" information processing stage.

Accordingly, the N400 has been used as an index of interpretation bias when providing individuals with emotional outcomes when reading scenarios. Participants from the general populations were induced by either positive or negative emotion, then they read sentences ended in either positive or negative ways (Chung et al., 1996). They found the interpretations generated by participants to the sentences, represented by N400 differences between positive and negative trials, were consistent with the induced emotions. Similarly, Moser and colleagues used N400 to investigate online interpretations in sub-clinical (Moser, Hajcak, Huppert, Foa, & Simons, 2008) and clinical populations (Moser, Huppert, Foa, & Simons, 2012). In these studies, participants listened to incomplete sentences (e.g., "You've just started reading a new book that you bought and you find it to be..."), and the ambiguity was resolved in a negative (e.g., "boring") or positive (e.g., "interesting") valenced manner by the final word presented on screen. Participants were instructed to decide whether or not the final word was a grammatically correct ending of the sentence. In their first study with high and low socially anxious participants (Moser et al., 2008), they did not find expected effects in N400, the valence differences in N400 between groups, or significant effects in reaction time. In a later study (Moser et al., 2012), however, a healthy comparison group showed a benign interpretation bias, evidenced by greater N400 amplitudes for negative words than benign words. In contrast, individuals with emotional disorders (e.g., social anxiety, major depressive disorder or dysthymia, and combined) showed the reverse pattern; their N400 amplitudes were greater for benign words compared to negative words, showing a negative interpretation bias. Consistent with the N400 results, reaction times showed that the healthy comparison group

reacted faster to benign words compared to negative words, indicating a benign interpretation bias that was lacking in the clinically anxious populations. The authors suggested that the inconsistencies between their studies may result from the materials being more relevant to clinical populations (2012 study) than the sub-clinical socially anxious participants (2008 study). Furthermore, they suggested that completing the questionnaires before the task may serve as an emotional trigger, which could lead to the stronger expectation violation effect (N400) as seen in the 2012 study. According to the Moser et al. (2012) study, N400 seems to be a promising index of online interpretation in clinical populations, but it remains unclear if it applies to high worriers in the general population. The current study used materials with a wide range of worry topics, making the materials relevant to high worriers. In addition, individuals' worry was activated before the task through a worry induction procedure designed to serve as an emotional trigger.

In sum, differences in interpretations made by high and low worriers (i.e., negative bias, no bias, benign bias), and at which stages of information processing (offline or online) such interpretation biases occur remains unclear, since previous research in high worriers has been limited to offline measures. Furthermore, the field more generally has only focused on either offline or online measures rather than assessing offline, online, and neurophysiological markers of interpretation within the same study. A crucial problem found in many interpretation measures is that they lack separate indices of benign and negative interpretation, to distinguish the presence of a negative bias from the absence of a benign bias. Relatedly, the relative bias within a group has not been explored, so conclusions can only be based on the comparisons between target groups and controls. The current study addresses these issues by investigating interpretation bias at different stages of information processing in high and low worriers using offline and online measures combined with EEG. The measures used in the present study included separate ratings for both benign and negative interpretations. This allowed

examination of interpretation bias within groups at different processing stages in order to elucidate how interpretation processes operate in worry.

This is the first study to examine interpretation bias in worry from both an early (online) information processing stage to a later (offline reflective) processing stage. Given that there are different methods in the field, we wanted to determine which paradigms are most sensitive to interpretation biases in high worriers. We included two online measures combined with EEG and two offline measures in the study to better capture the characteristics of interpretation bias in high and low worriers. Based on the previous studies that included independent ratings and compared negative and benign interpretations within groups (e.g., Eysenck et al., 1991; Hirsch & Mathews, 1997; 2000; Moser at al., 2012), high worriers were hypothesised to show no signs of interpretative biases (i.e. they generated both benign and negative interpretations equally), both in offline and online measures, whereas low worriers were hypothesized to show signs of benign interpretation bias and it was not clear whether both offline and online measures would capture this bias.

## Methods

### Design

This study aimed to understand the relationship between interpretation bias and worry using multiple methods that capture the different processing stages of interpretation bias. A between-subjects design was employed with group as the between-subjects factor. Individuals with high and low levels of worry were recruited to compare the interpretation bias between these two groups. Interpretation bias was assessed by two offline and two online tasks that combined with ERP measures. A period of worry was experimentally induced before each online task, to activate the interpretation biases. Self-report ratings (offline tasks), reaction time and ERP amplitudes (online tasks) were used as outcome measures of

interpretation bias. The levels of worry between groups were also examined to establish group differences. Self-report questionnaires were administered to assess the extent of worry. In addition, one filler task was conducted before the offline task to eliminate carryover effects from online tasks to offline task. The study was approved by King's College London Research Ethics Committee, and all participants provided written informed consent.

**Participants**

Participants were recruited through online advertisements at King's College London and throughout South East London. The *Penn State Worry Questionnaire* (PSWQ) was used as the screening questionnaire to identify participants' worry levels. The cut-off score was 56 or more to identify high worriers because this score is one deviation below the mean of diagnosed GAD (Molina, & Borkovec, 1994). For low worriers, the cut-off score was 39 or less, because it is 0.5 standard deviations below the mean of non-anxious populations (Molina, & Borkovec, 1994). Participants were excluded if they fell between 40 and 55 on PSWQ, were not fluent in English, had vision that was not normal or correct-to-normal, or who had a seizure disorder or current brain injury. Based on these criteria, 28 high worry participants and 27 low worry participants who met the criteria were included in this study. Sixty percent of participants were female and the mean age was 27.69 years (SD = 9.47). Groups did not differ significantly on age (high worry: M = 26.07, SD = 8.90; low worry: M = 29.37, SD = 9.92; t (53) = -1.30, $p$ = .199). Although non-significant, there were more female participants in the high worry group than low worry group (71.43% female in high worry group, 48.15% female in low worry group; $\chi^2(1) = 3.10$, $p$ =.078). Three high worriers and one low worrier were not native English speakers but self-reported as completely fluent

in English (5/5)[1]. All participants except 6 low worriers were right-handed according to self-report. They completed a battery of self-report questionnaires after they enrolled in the study. The high worry group reported significantly higher levels of worry (PSWQ), anxiety (GAD-7), worry in five domains (WDQ), and depression (PHQ-9) than the low worry group, all *p*s < .001 (see Table. 1), as expected. See below for the measures section with details of questionnaires.

## Measures

### Questionnaires.

***Penn State Worry Questionnaire (PSWQ)***. The PSWQ (Meyer, Miller, Metzger, & Borkovec, 1990) includes 16 items and measures trait worry (e.g., "Many situations make me worry"), using a 5-point scale from 1 (*not at all typical of me*) to 5 (*very typical of me*). Therefore, the summed scores range from 16 to 80 with 5 reverse-scored items. A higher score indicates a higher level of worry. The PSWQ has good psychometric properties, internal consistency, short-term test-retest reliability and convergent and criterion-related validity (Brown, Antony, & Barlow, 1992; Davey, 1993). Cronbach's alpha in this study was .97.

***Worry Domains Questionnaires (WDQ)***. The WDQ (Tallis, Eysenck, & Mathews, 1992) contains 25 items relating to five worry domains to assess the worry level of specific

[1] Given that language fluency may affect the N400 effect (Martin et al., 2013), we compared N400 analyses with and without these four non-native speakers and found the results were identical to each other. Therefore, we presented the N400 results with these four participants in the results section.

content (e.g., I worry that I will lose close friend) with a 5 points scale from 1 (*Not at all*) to 5 (*Extremely*). The five domains are: lack of confidence, aimless future, work performance, and finances. The summed scores for all items range from 21 to 105, a higher score indicates higher worry level. The WDQ also has good internal consistency and test-retest correlation (Stöber, 1998). Cronbach's alpha in this study was .97.

***Generalized Anxiety Disorder 7-item scale (GAD-7).*** The GAD-7 (Spitzer, Kroenke, Williams, & Löwe, 2006) is a self-report scale to measure anxiety severity and identify probable individuals with GAD and has 7 items. It inquiries about the frequency of symptoms, scored from 0 (*Not at all*) to 3 (*Nearly every day*), in the past two weeks. For example, "Feeling nervous, anxious, or on edge". The summed scores range from 0 to 21; a higher score indicates higher anxiety level. It has good internal consistency, test-retest reliability, diagnostic criterion validity in the patient sample (Spitzer et al., 2006), and is also a reliable and valid self-report measure for anxiety in the general population (Löwe et al., 2008). Cronbach's alpha in this study was .95.

***Patient Health Questionnaire (PHQ-9)***. The PHQ-9 (Kroenke, Spitzer, & Williams, 2001) is a self-report questionnaire, which rates the frequency of depressive symptoms over the past 2 weeks. It has 9 items, scored from 0 (*Not at all*) to 3 (*Nearly every day*). The summed scores range from 0 to 27; higher score means more severe depressive symptoms. The internal reliability, test-retest reliability, and diagnostic criterion validity are all good in the PHQ-9 (Kroenke et al., 2001). Cronbach's alpha in this study was .92.

**Offline interpretation bias measures.**

***Scenario task.*** The scenario task was developed by Butler and Mathews (1983) and is a widely used task (e.g., Amir et al., 1998; Stopa & Clark, 2000). In this study, 16 brief scenarios with content related to a range of worry domains were used. The scenarios were

adapted from Krahé, Mathews, Whyte, & Hirsch (2016) and other studies, including Mathews & Mackintosh (2000); Hirsch at al. (2009); Hayes et al. (2010); and Grol et al., (2018). Each scenario remained ambiguous and three possible interpretations were displayed randomly: one was a positive interpretation, one was a negative interpretation and the third was a neutral interpretation (see Table 2 for an example). Participants were asked to read the scenario first and then rate the likelihood with which each interpretation sentence would come to their mind, using a nine-point Likert-type scale (1 very unlikely to 9 very likely). The likelihood ratings for each interpretation type were averaged to provide a mean score. Greater likelihood ratings for negative targets indicated a greater negative interpretation bias; likewise, greater ratings for positive interpretation targets indicated a greater positive interpretation bias.

***Recognition task.*** The recognition task (Eysenck et al., 1991) is widely used (e.g., Mathews & Mackintosh, 2000; Murphy, Hirsch, Mathews, Smith, & Clark, 2007) and comprises 16 scenarios with descriptive titles. The scenarios were adapted from the same sources as the scenario task, but the content did not overlap with the scenario task. The scenarios were related to worry and ambiguity remained unresolved. Participants were asked to read the scenarios first, make a fragment completion judgments of the final word, then complete a comprehension question. After they read all scenarios, a list of possible interpretations with a title for each scenario were randomly presented, which contained a positive target interpretation, a negative target interpretation, a positive foil and a negative foil. Foils were not the interpretations of the ambiguity in the scenario and were used to assess a general valence effect (see Table 2 for an example). Participants were asked to rate how similar each sentence was to the original scenario using a four-point Likert-type scale (1 very different in meaning - 4 very similar in meaning). The scores of each sentence type were averaged. The

greater the similarity ratings for negative interpretations, the greater the negative interpretation bias. Likewise, greater similarity ratings for positive interpretations indicated greater positive interpretation bias.

**Online interpretation bias measures.**

***Material development and piloting.*** The materials used in online measurements were worry-related sentences, which were created by the authors and adapted from previous studies (Hirsch et al., 2018). For example: "You cancel the lunch plan with yourcousin due to your schedule, and you know he will be____"; "Your teacher provides you with a lot of feedback, most of it is_____". The sentences were piloted online in the general population using Amazon Mechanical Turk to ensure that both the benign and negative concepts were generated. The cloze-probabilities of the target words based on the pilot did not differ between two tasks[2]. Post study analysis of differences between conditions in word frequency (based on the Corpus of Contemporary American English (COCA), https://www.english-corpora.org/coca/) showed that the benign target word frequencies were larger than negative target word frequencies in each task[3]. This may affect the response latency as higher frequency words are generally processed faster than lower frequency words (Monsell, 1991). Therefore, participants may have potentially faster response to benign words than negative words due to the frequency effect. However, given that the aim of the study was to investigate

[2] In the final materials sets, 64 word trial sentences in the LDT and 63 sentences in the SWAT (out of 80 word trials or related-word trials) were selected from the pilot materials. For the benign trials, the mean cloze-probability was 22.33% (*SD*: 19.58%) in the LDT and 24.32% (*SD*: 20.97%) in the SWAT ($t(125) = -.55$, $p = .581$). For the negative trials, the mean cloze-probability was 14.66% (*SD*: 12.37%) in the LDT and 16.92% (*SD*: 17.63%) in the SWAT ($t(125) = -.84$, $p = .403$).

[3] The average logarithm of frequency (LogFQ) in the LDT was 9.71 (*SD* = 1.64) for negative trial words, and 10.52 (*SD* = 1.63) for benign trial words. The benign trials words LogFQ was larger than negative trial words LogFQ ($t(158) = -3.13$, $p = .002$). The average LogFQ in the SWAT was 9.61 (*SD* = 1.62) for negative trial words, and 10.41 (*SD* = 1.36) for benign trial words, benign was larger than negative trials words LogFQ ($t(158) = -3.37$, $p = .001$).

differences in interpretation bias between high and low worriers, we provided the same materials to high and low worry groups to see how they responded to those materials. We expected since every participant was exposed to the same lexico-semantic characteristics, the differences in the responses between groups represented the difference in interpretations.

***Worry phase.*** Before the online tasks, participants were asked to think about a current worry subject that was relevant to themselves and to worry about it for four minutes. This procedure was to ensure that their worry was activated and could affect the process of judgment (Hayes, Hirsch, & Mathews, 2008). Questions were asked to facilitate the identifying worrying thoughts, for example, "What's worrying about [worry topic]?", "What do you fear might happen?" The self-reported anxiety levels of both groups after both worry phases were checked, all of them were higher than the anxiety levels at the baseline (the first rating before experiment) and levels of anxiety in both phases did not differ from each other[4].

***Lexical decision task (LDT).*** The LDT (Hirsch & Mathews, 1997, 2000) has been used to assess interpretation bias (e.g., Bisson & Sears, 2007). In this study, the LDT included 120 short sentences related to worry topics. Every sentence was only displayed once and paired with one letter string. Within the 120 trials, 40 were paired with benign words, 40 were presented with negative words and 40 with non-words. There were two sets of materials, set A and B, that were counterbalanced across participants. In the sentences paired with words, 40/80 sentences were completed with words that matched the benign interpretation in Set A of materials. In set B, the same 40 sentences were completed with the words that matched the negative interpretation, and vice versa for other 40 sentences. The 40 non-words trials

4 The high worry group had higher levels of anxiety in phase 1 and 2 compared with baseline (phase one: $t$ (27)= -3.38, $p$ = .002; phase two: $t$ (27) = -2.26, $p$ = .032), and both phases did not differ from each other ($t$ (27) = 1.35, $p$ = .188). The results were the same with the low worry group, in which higher levels of anxiety in phase 1 and 2 compared with baseline were found (phase one: $t$ (26)= -3.75, $p$ = .001; phase two: $t$ (26) = -3.30, $p$ = .003), and the levels of anxiety in both phases did not differ from each other ($t$ (26) = 0.71, $p$ = .487).

were the same for both sets (see Table 2 for examples).The three types of letter strings were allocated equally to two blocks. For each trial, a fixation cross first appeared for 200ms and was then followed by an incomplete sentence (see Table 2 for an example). Participants were instructed to press a button after they had read the sentence. Then, a blank screen replaced the sentence for 500ms to 750ms, followed by a benign or negative final word that resolved the ambiguity, or a non-word final word until response. Participants were asked to decide whether the words were real words or not. They were required to press a key with their left index finger if it was a word and press a key with their right index finger if it was not a word. Then 50% of the trials were followed by a comprehension question to ensure participants read the sentences correctly. All sentences and target words appeared in white, 18 pixels, against a black background.

The reaction time medians[5] and N400 amplitudes of benign or negative words were computed as interpretation bias indices. The faster reaction time median of the negative words indicated a more negative interpretation bias; the faster reaction time median of the benign words indicated a more benign interpretation. The larger negative N400 amplitude (more negative value of the amplitude) following the words indicated that the words were not consistent with participant's expectations (see below for details). Therefore, the larger negative N400 amplitude for benign words meant greater negative interpretation bias.[6]

***Sentence word association task (SWAT).*** The SWAT was adapted from the "Word Sentence

[5] The reaction time medians were used in this study instead of means because reaction time means are positively *skewed and medians are* much more insensitive to the skew of the distribution (Baayen & Milin, 2010) and in keeping with previous research using the LDT task (Hirsch & Mathews, 1997, 2000; Hirsch et al., 2003; Hirsch et al., 2006). Analysis of reaction time mean data is presented in the supplementary materials and shows the same pattern of results as the median reaction time data.

[6] The N400 mean amplitude of non-words was also computed to examine the expectation violation effect compared to real word targets.

Association Task" (Beard & Amir, 2009) but in the present study, the sentence was presented before the word (rather than word followed by sentence) because it seems more ecologically valid in terms of encountering an ambiguous situation first and then generating a meaning of that situation. This adapted form has already been used in another study (Sears, Bisson & Nielsen, 2011). There were two sets of materials in this study, set A and B, which were counterbalanced across participants. One hundred and twenty short sentences related to worry topics were in each set of materials. In the 120 trials, 40 words presented were related-benign words, 40 were related-negative and 40 were non-related words. In the sentences paired with related words, 40/80 sentences paired with related-benign words in Set A of material were paired with related-negative words in Set B, and vice versa for the 40 related-negative words trials. The 40 sentences that followed non-related words were the same in both sets (see Table 2 for examples). Participants were asked to decide whether or not the sentence and word were related. They were required to press a key with their left index finger if the word was related to the sentence or press a key with their right index finger if they were not related. Different from the previous studies, 40 non-related-word trials were included to ensure participant did not endorse all the trials or respond randomly. Comprehension questions were also included in our study in 50% of trials to ensure participants read the sentence correctly.

There were three indices of interpretation bias in this task. As in previous studies (Beard & Amir, 2009), the greater percentage of endorsing the negative words indicated a greater negative interpretation bias and vice versa for benign words. Shorter reaction time medians to endorse negative words also indicated a greater negative interpretation bias. As in the LDT, a smaller N400 amplitude of negative words indicated greater negative interpretation bias. The N400 mean amplitude of non-related words was also computed to examine the expectation violation effect compared to the related target.

**Filler task.** The filler task was used to eliminate any potential carryover effects from the previous task and to reduce possible group differences in mood[7]. The "speed of comprehension task" (Baddeley, Emslie, & Nimmo-Smith, 1992) was used in this study, which asked participants to judge whether sentences were true or false (e.g., "Beer lives in trees"). Participants were instructed that speed was not important and did the task for two minutes.

**Electroencephalography (EEG) Recording and Data Processing Procedures**

While participants completed the online interpretation bias measures, EEG was recorded continuously using NuAmp amplifier (Neuroscan Inc.) with 1000 Hz sampling rate, recorded from 32 Ag/AgCl electrodes (FP1/FP2, F3/F4, F7/F8, FC3/FC4, FT7/FT8, C3/C4, T3/T4, CP3/CP4, TP7/TP8, P3/P4, T5/T6, O1/O2, Fz, FCz, Cz, CPz, Pz, and Oz) placed on the scalp with an elasticated cap, positioned according to the 10-20 international system (AEEGS, 1991). No online filter was applied during recording. Vertical eye movements were recorded using electrodes placed on the supraorbital and infraorbital ridges of the left eye [vertical electro-oculogram (VEOG)], and horizontal eye movements by electrodes placed on the outer canthi of the right and left eyes [horizontal electro-oculogram (HEOG)]. Additional electrodes were used as ground and reference sites. Electrodes were referenced to the right mastoid site (A2) during recording. The electrode between FPz and Fz (AFz) on the midline was served as the ground electrode. Impedances were kept below 10kΩ.

[7] The mood ratings after the filler tasks showed no differences in the levels of anxiety and depression between groups when the baseline mood ratings (the first rating before experiment) were included as a covariate variable. The low worry group showed a higher level of happiness than the high worry group after the first filler task. Group main effects of mood ratings after the first filler task: anxiety, $F(1, 52) = 3.55$, $p = .065$, $\eta^2_p = .06$; depression, $F(1, 52) = 0.14$, $p = .713$, $\eta^2_p = .00$; happiness, $F(1, 52) = 6.28$, $p = .015$, $\eta^2_p = .108$. Group main effects of mood ratings after the second filler task: anxiety, $F(1, 52) = .30$, $p = .587$, $\eta^2_p = .00$; depression, $F(1, 52) = 0.02$, $p = .891$, $\eta^2_p = .00$; happiness, $F(1, 52) = 3.05$, $p = .087$, $\eta^2_p = .06$.

Signal processing was performed using MATLAB software with EEGLAB (Delorme & Makeig, 2004) and ERPLAB toolboxes (Lopez-Calderon & Luck, 2014). Data were re-referenced to an average of both mastoid sites (A1 & A2), and filtered using a 30 Hz low-pass filter and a 0.01 Hz high-pass Butterworth filter with 12 dB/octave roll-off. Independent component analysis (ICA) was conducted on the continuous EEG. The eye-artifact (eye movement and blink) components derived from the ICA were discarded from the EEG data. Then, the EEG data were segmented for each trial beginning 200ms prior the onset of the target words and continuing for 800ms after target words onset (1000ms total duration). Baseline correction was conducted using the 200ms before the onset of the targets. Trials with incorrect responses were rejected. Additional artifacts were rejected upon visual inspection on a trial-by-trial basis. Then the ERP data were averaged separately according to the target word type. The N400 mean amplitudes were measured at 6 centro-parietal channels (Cz, C3, C4, Pz, P3, and P4) with a 300ms to 450ms time window. Noisy data in which an ERP was not apparent and data with too few trials (less than 70% in one experimental condition) were excluded from the data analysis. Group mean amplitudes were then computed for each word type.

In the final analysis, the average of rejected IC (independent component) in the LDT was 2.33. The average of accepted trials was 82.11% in benign trials and 82.22% in negative trials, the accepted trials numbers did not differ from each other ($t$ (44) = -0.09, $p$ = .928). In the SWAT, the average of rejected IC was 1.96, the average of accepted trials was 97.17% in benign trials and 94.44% in negative trials. The accepted benign trials were more than the accepted negative trials ($t$ (44) = 3.08, $p$ = .004)[8]

[8] The data is based on the final data that has excluded participants who did not meet criteria. See "Data Preparation for Online and Offline Measures of Interpretation Bias" part for details. The benign and negative trial

**N400 Component**

In this study, N400 amplitude was measured in the 300-450ms time window after target word onset at the Cz, C3, C4, Pz, P3, and P4 electrode sites. N400 reflects the ease with which information is integrated into the context based on individuals' semantic memory (Kutas & Federmeier, 2000). Therefore, information that is harder to integrate into a former context, or that violates one's expectation of the context, generates a larger negative N400 amplitude. N400 has previously been used as an index of interpretation bias (e.g., Moser et al., 2008, 2012). In particular, if the given interpretation of the context is not consistent with an individual's expectation, a larger N400 amplitude will be elicited compared to the interpretation consistent with the context. In this study, the context was the ambiguous scenarios and was followed by their inferences (target words).

**Experimental Procedure**

Twenty-four hours before the experimental session, a link to the questionnaires and scenario task were sent to participants. They were asked to complete the online questionnaires and scenario task before they came to the session. After participants arrived at the session, they first gave informed consent and the EEG cap was fitted. Then they completed one worry phase, followed by the online interpretation bias measure (LDT or SWAT, counterbalanced across participants) with EEG recorded continuously. Then another worry phase and the remaining online interpretation bias measure were administrated with EEG. Following this, the EEG cap was removed. After completing the filler task, the recognition task was done. Finally, participants were debriefed and compensated for their time. The session lasted about 2.5 hours

numbers' difference in the SWAT was only 1 trial. We did not expect it would affect further analysis based on 1 trial difference.

(see Figure 1 for the study flow chart).

**Data Preparation for Online and Offline Measures of Interpretation Bias**

In order to examine if participants completed the tasks correctly, the response accuracies for the comprehension questions were first analysed. The mean accuracy in the recognition task was 77.61% (SD = 8.84%). For the LDT, the mean accuracy was 86.27% (SD = 4.26%). For the SWAT, the mean accuracy was 87.00% (SD = 6.69%). No significant differences in accuracy between the two groups were found in these three tasks[9]. In the LDT, the response accuracies for the word/non-word judgments were also examined. The mean accuracy was 97.18% (SD = 2.83%) across all participants. There was a significant difference between groups, high worry: 95.95% (SD = 4.84%) vs. low worry group: 98.45% (SD = 3.04%), t (53) = -2.30, $p$ = .025.[10]

Across tasks, participants whose percentage of correct completions of the comprehension questions and the word/non-word judgments was more than 2.5 standard deviations below mean were excluded. One low worrier on the recognition task and two high worriers in the

[9] T-tests for accuracies of comprehension questions between groups: Recognition task, $t$(53) = -1.10, $p$ = .278; LDT, $t$(53) = -1.77, $p$ = .081; SWAT, $t$(53) = -1.40, $p$ = .168.

[10] The results for the major hypothesis in LDT did not change when accuracy of word/non-word judgment was used as a covariate in analyses. Reaction time median, Group: $F$(1, 47) = .13, $p$ = .725, $\eta^2_p$= .00; Word type: $F$(1, 47) = 4.90, $p$ = .032, $\eta^2_p$= .09 (benign<negative); Group x Word type: $F$(1, 47) = .01, $p$ = .913, $\eta^2_p$= .00. N400 amplitude, *Group*: $F$(1, 43) = 0.06, $p$ = .810, $\eta^2_p$< .01; *Valence*: $F$(1, 42) = 1.97, $p$ = .168, $\eta^2_p$= .05; *Electrode Sites*: $F$(5, 210) = 1.4, $p$ = .226, $\eta^2_p$= .03; *Group* x *Valence* interaction ($F$(1, 42) = 6.00, $p$ = .019, $\eta^2_p$= .13); Group x *Electrode Sites*: $F$(5, 210) = 0.86, $p$ = .51, $\eta^2_p$= .05; *Valence x Electrode Sites*: $F$(5, 210) = 0.60, $p$ = .703, $\eta^2_p$< .01; of *Group* x *Valence* x *Electrode Sites*: $F$(5, 210) = 1.00 , $p$ = .417, $\eta^2_p$= .02. Therefore, we only presented data without the covariate variable.

LDT were excluded from the data analysis for poor performance on the comprehension questions. Furthermore, one high worrier and two low worriers were excluded because of poor accuracy on the word/non-word judgments. In the SWAT, two high worriers were excluded because of poor performance on the comprehension questions. For the EEG analysis, two low worriers in the LDT was excluded due to an insufficient number of trials after inspecting the data for artefacts. Furthermore, two high worriers and one low worrier in the LDT, four high worriers and three low worriers in the SWAT were excluded due to noisy data in which an ERP was not apparent. In the SWAT, another low worrier was excluded due to recording failure. See Table. 3 for the final participant numbers in each task. The N400 calculation for the LDT was based on lexical judgment correct trials. For the SWAT, the N400 calculation was based on all the endorsed trials (i.e., categorised by the participant as related) since there was no explicit right or wrong for relatedness judgement.

**Analytic Strategy**

Analyses of self-report, behavioural, and ERP measures proceeded as follows:

To examine the group differences on worry levels and psychological distress, t-tests were used to compare questionnaire scores between groups.

There were four interpretation bias measures. Performance on the comprehension questions and judgments between groups in these measurements were tested by t-tests before further analysis. For interpretation bias in the offline measures, the likelihood rating in the scenario task was the dependent variable, tested by a 2 x 3 mixed ANOVA with Group (high worry vs. low worry) as a between-group variable and Valence (positive, negative, and neutral)

as a within-group variable. In the recognition task, a 2 x 2 x 2 ANOVA was conducted on mean similarity rating with Group (high worry vs. low worry) as a between-group variable, Sentence type (target vs. foil) and Valence (negative vs. positive) as a within-group variables. For the online measures, the reaction time median in the LDT was tested by a 2 x 2 mixed ANOVA, with Group (high worry vs. low worry) as a between-group variable and Valence (benign vs. negative) as a within-group variable. In the SWAT, endorsement rate and reaction time median were tested by 2 x 2 mixed ANOVAs, with Group (high worry vs. low worry) as a between-group variable and Valence (benign vs. negative) as a within-group variable.

For the ERP analysis in both the LDT and SWAT, we first examined the general N400 expectation violation effect and then conducted the test for interpretation bias. In particular, we expected that the non-words or non-related words would violate participants' expectation because they would expect to see real words or related words following the sentences. Therefore, non-words or non-related words were expected to elicit greater N400 amplitudes than the real words or related words. To examine this expectation violation effect, the N400 amplitudes for the words and non-words (or related words vs. non-related words) were compared by a 2 (Word Type) x 6 (Electrode Site) repeated measures ANOVA. Having established the expectation violation effect, the main analysis for interpretation bias was conducted. Three-way (2 x 2 x 6) ANOVAs were conducted for 6 centro-parietal electrode sites with Group (high worry vs. low worry) as a between-group variable and Valence (benign vs. negative) and Electrode Site (C3, Cz, C4, P3, Pz, and P4) as within-group variables.

## Results

### Offline Measures of Interpretation Bias

**Scenario task.** The 2 x 3 mixed ANOVA conducted on the mean likelihood rating for the different interpretations showed no significant main effect of Group, $F$ (1, 53) = .12, $p$ = .732,

but a significant main effect for Valence, $F$ (2, 106) = 19.19, $p$ < .001, $\eta^2_p$ = .27. Bonferroni-adjusted follow-up tests indicated higher likelihood rating for the positive and neutral than negative outcomes across groups ($p$s < .001). However, the likelihood ratings for positive and neutral interpretations did not differ from each other ($p$ = 1.00). Importantly for our hypothesis, this main effect was qualified by a significant Group by Valence interaction, $F$ (2, 106) = 21.49, $p$ < .001, $\eta^2_p$ = .29. In line with our hypothesis, direct comparisons of ratings revealed that the low worry group rated positive and neutral outcomes as more likely to come to their mind than negative outcomes, positive vs. negative: $t$ (26) = 6.69, $p$ < .001, $d$ = 2.11; neutral vs. negative: $t$ (26) = 7.26, $p$ < .001, $d$= 2.00; positive vs. neutral: $t$ (26) = 1.85, $p$ = .075. However, there were no differences between interpretations in high worry group, positive vs. negative: $t$ (27) = -0.43, $p$ = .670; positive vs. neutral: $t$ (27) = -1.69, $p$ = .103; neutral vs. negative: $t$ (27) = 0.22, $p$ = .830 (see Table 3 and Figure 2). In summary, the results in scenario task were consistent with the hypothesis that the low worry group was more prone to making benign interpretations than negative interpretations. However, the high worry group did not show any significant differences when rating benign and negative interpretations.

**Recognition task.** A three-way ANOVA was conducted on similarity ratings. The main effects of Sentence type and Valence were significant, Sentence type: $F$ (1, 52) = 140.85, $p$ < .001, $\eta^2_p$ = .73; Valence: $F$ (1, 52) = 10.82, $p$ = .002, $\eta^2_p$ = .17. They revealed higher ratings for the target sentences compared to the foil sentences, and higher ratings for the positive interpretations than the negative interpretations. The main effects were qualified by a significant Group x Valence interaction, $F$ (1, 52) = 13.97, p < .001, $\eta^2_p$ = .21, Sentence

type x Valence, $F$ (1, 52) = 11.72, $p$ = .001, $\eta^2_p$ = .18, and Group x Sentence type x Valence interactions, $F$ (1, 52) = 4.05, $p$ = .049, $\eta^2_p$ = .07.

To understand the three-way interaction, we looked at the Sentence Type x Valence interaction for each Group separately. For the high worry group, the Valence main effect and interaction were non-significant, Valence: $F$ (1, 27) = 0.07, p = .787, $\eta^2_p$ < .01, Valence x Sentence type: $F$ (1, 27) = 0.82, p = .372, $\eta^2_p$ = .03, while there was a significant Sentence type main effect, $F$ (1, 27) = 93.64, p < .001, $\eta^2_p$ = .78, showing that the rating for target sentences was higher than the rating for foil sentences. For the low worry group, there was a significant Sentence type main effect, $F$ (1, 25) = 53.62, p < .001, $\eta^2_p$ = .68, showing that the rating for target sentences was higher than the rating for foil sentences. The Valence main effect was also significant, $F$ (1, 25) = 42.04, $p$ < .001, $\eta^2_p$ = .63, showing that the rating for positive sentences was higher than the rating for the negative sentences. The main effects were qualified by a significant Sentence type by Valence interaction, $F$ (1, 25) = 19.70, p < .001, $\eta^2_p$ = .44. Direct comparisons showed that the low worry group rated positive target sentences higher than negative target sentences, $t$ (25) = 6.83, $p$ < .001, $d$= 1.23. They also rated positive foils higher than negative foils, $t$ (25) = 4.63, $p$ < .001, $d$= 0.67 (see Table 3 and Figure 2). In summary, the ratings in recognition task indicated that high worriers lacked a benign interpretation bias displayed by the low worry group.

## Online Measures of Interpretation Bias

**Lexical decision task (LDT).** In the LDT, the reaction time median and the N400 amplitude following the words were the indices of interpretation bias.

***Reaction time median.*** A 2 x 2 mixed ANOVA on reaction time medians showed a significant main effect of Valence, $F$ (1, 48) = 14.82, $p$ < .001, $\eta^2_p$ = .24, demonstrating a faster reaction time to benign valence trials than negative valence trials (see Table 3). The

main effect of Group and interaction of Group x Valence were not significant, Group: $F$ (1, 48) = .10, $p$ = .750, $\eta^2_p$ < .01; Group x Valence: $F$ (1, 48) = 1.84, $p$ = .181, $\eta^2_p$ = .04. However, post-hoc tests based on a priori hypotheses revealed that low worriers reacted faster to benign compared to negative valence trials, $t$ (24) = 4.12, $p$ < .001, $d$ = 0.45, while high worriers failed to demonstrate a significant difference between two valence trials, $t$ (24) = 1.61, $p$ = .121, $d$ = 0.15. The results revealed the predicted benign interpretation bias in low worry group.

***N400 amplitude.***[11] The 2 x 2 x 6 mixed ANOVA performed on N400 amplitude at six electrode sites. There was a significant main effect of *Electrode Sites* ($F$ (5, 215) = 19.56, $p$ < .001, $\eta^2_p$ = .31), and a significant *Group* x *Electrode Sites* interaction[12], $F$ (5, 215) = 2.44, $p$ = .036, $\eta^2_p$ = .05. There were no significant *Group* or *Valence* main effects, nor *Valence* x *Electrode Sites* interaction in this analysis (Group: $F$ (1, 43) = 0.06, $p$ = .810, $\eta^2_p$ < .01; *Valence*: $F$ (1, 43) = .81, $p$ = .372, $\eta^2_p$ = .02; *Valence* x *Electrode Sites*: $F$ (5, 215) = 0.33, $p$ = .895, $\eta^2_p$ < .01). Furthermore, there was no significant three-way interaction of *Group* x *Valence* x *Electrode Sites*: $F$ (5, 215) = 1.15, $p$ = .338, $\eta^2_p$ = .03). Importantly for our hypotheses, there was a significant *Group* x *Valence* interaction ($F$ (1, 43) = 4.10, $p$ = .049, $\eta^2_p$ = .09), regardless of *Electrode Sites*. Post-hoc comparisons showed that the low

[11] In order to examine the N400 effect of expectation violation, a repeated measures ANOVA with Word Type (word vs. non-word) and Electrode Site (Cz, C3, C4, Pz, P3 and P4) as within group variables was first conducted. In line with our prediction, there was an enhanced N400 amplitude for non-words compared to real words, $F$(1, 44) = 18.97, $p$ < .001, $\eta^2_p$ = .30, indicating a valid expectation violation effect. Therefore, further analyses for interpretation bias were continued.

[12] For the *Electrode Sites* main effect, post-hoc comparisons showed C3 had a greater amplitude than the other five electrode sites (*ps* < .004). Pz had a smaller amplitude than C3, Cz, C4, and P3 electrode site (*ps* < .002). P3 had a greater amplitude than P4 ($p$ = .017). For the *Group* x *Electrode Sites* interaction, post-hoc comparisons showed no significant differences between groups were found in any of the electrode sites (*ps* > .324).

worry group had a significantly greater negative amplitude following negative valence trials than benign valence trials ($p$ = .047), while the amplitudes of valence trials were not significantly different in the high worry group ($p$ = .427)[13] (see Table 3. and Figure 3).

**Sentence word association task (SWAT).** In the SWAT, three indices were used to measure interpretation bias, namely endorsement rate, reaction time median, and N400 amplitude.

***Endorsement rate.*** No main effects were found in the 2 x 2 mixed ANOVA, Group: $F$ (1, 51) = .29, $p$ = .594, $\eta^2_p$= .01; Valence: $F$ (1, 51) =3.06, $p$ = .086, $\eta^2_p$= .06. However, in line with our hypothesis, the interaction of Group x Valence was significant, $F$ (1, 51) = 13.81, $p$ = .001, $\eta^2_p$= .21. Follow up tests showed that only the low worry group had a greater endorsement rate for benign interpretations compared to negative interpretations, $t$ (26) = -4.32, $p$ < .001, $d$ = -.95. No difference was found in the high worry group between negative and benign endorsement rates, $t$ (25) = 1.26, $p$ = .219, $d$ = 0.32 (see Table 3). The results supported our hypothesis that the benign interpretation bias was only found in the low worry group but not in the high worry group, where they lack of benign interpretation bias.

***Reaction time median***. The mixed ANOVA was performed on reaction time medians for

[13] The results did not meaningfully differ when the excluded data (two high worriers and one low worrier) were included in the analysis presented here. *Group*: $F$(1, 46) = 0.22, $p$ = .643, $\eta^2_p$< .01; *Valence*: $F$(1, 46) = .97, $p$ = .331, $\eta^2_p$= .02; *Electrode Sites*: $F$(5, 230) = 19.79, $p$ < .001, $\eta^2_p$= .30; Group x *Electrode Sites*: $F$(5, 230) = 2.48, $p$ = .032, $\eta^2_p$= .05; *Valence x Electrode Sites*: $F$(5, 230) = 0.27, $p$ = .929, $\eta^2_p$< .01; of *Group* x *Valence* x *Electrode Sites*: $F$(5, 230) = 1.09, $p$ = .365, $\eta^2_p$= .02. A significant *Group* x *Valence* interaction ($F$(1, 46) = 5.23, $p$ = .027, $\eta^2_p$= .10) was shown. Post-hoc comparisons showed that the low worry group had a significantly greater negative amplitude following negative valence trials than benign valence trials ($p$ = .028), while the amplitudes of valence trials were not significantly different in the high worry group ($p$ = .351).

trials participants endorsed. No significant main effects were found in this analysis, Group: $F$ (1, 51) = .08, $p$ = .776, $\eta^2_p$ < .01; Valence: $F$ (1, 51) = 2.41, $p$ = .127, $\eta^2_p$ = .05. However, there was a significant interaction of Group x Valence, $F$ (1, 51) = 8.44, $p$ = .005, $\eta^2_p$ = .14. Follow-up tests showed that the low worry group was faster to endorse benign compared to negative interpretations, $t$ (26) = 3.53, $p$ = .002, $d$= .41 (see Table 3). However, the high worry group showed no difference between benign and negative trials ($p$ = .395). In summary, our hypotheses were supported that only the low worry group showed an interpretation bias toward benign interpretations and high worry group lack of this benign bias.

***N400 amplitude*[14]**. The 2 x 2 x 6 mixed ANOVA performed on N400 amplitude showed a significant main effect of *Electrode Sites* ($F$ (5, 215) = 18.59, $p$ < .001, $\eta^2_p$ = .30)[15]. However, inconsistent with our hypotheses, no significant *Group* x *Valence* interaction was found ($F$ (1, 43) = 0.29, $p$ = .596, $\eta^2_p$ = .01). No other significant main effects or interactions were found in this analysis (see Table 3. and Figure 4), *Group*: $F$ (1, 43) < 0.01, $p$ = .967, $\eta^2_p$ < .01; *Valence*: $F$ (1, 43) = .26, $p$ = .616, $\eta^2_p$ = .01; *Group* x *Electrode Sites*: $F$ (5, 215) = 0.22, $p$ = .953, $\eta^2_p$ = .01; *Valence* x *Electrode Sites*: $F$ (5, 215) = 0.54 , $p$ = .748, $\eta^2_p$ = .01; *Group* x *Valence* x *Electrode Sites*: $F$ (5, 215) = 0.91 , $p$ = .347, $\eta^2_p$ = .02.[16]

---

[14] The expectation violation effect was also examined in the SWAT. A repeated measures ANOVA on the N400 amplitudes showed greater amplitude for non-related words than related words, $F$(1, 44) = 125.13, $p$ < .001, $\eta^2_p$ = .74, supporting the N400 expectation violation effect. Therefore, further analysis was performed only on benign and negative words.

[15] Post-hoc comparisons for the *Electrode Sites* main effect showed C3 had a greater amplitude than Cz, P3, Pz, and P4 electrode site (*ps* < .016). Pz had a smaller amplitude than the other five electrode sites (*ps* < .033). P3 had a smaller amplitude than C3, Pz, and P4 (*ps* < .046).

[16] The results did not meaningfully differ when the excluded data (four high worriers and three low worriers) were included in the analysis presented here. *Group*: $F$(1, 50) = 0.01, $p$ = .970, $\eta^2_p$ < .01; *Valence*: $F$(1, 50) = .37, $p$ = .548, $\eta^2_p$ = .01; *Electrode Sites* ($F$(5, 250) = 22.25, $p$ < .001, $\eta^2_p$ = .31; *Group* x *Electrode Sites*: $F$(5, 250) = 0.13, $p$ = .986, $\eta^2_p$ = .01); *Valence x Electrode Sites*: $F$(5, 250) = 0.58 , $p$ = .715, $\eta^2_p$ = .01; *Group* x *Valence* x *Electrode Sites*: $F$(5, 250) = 0.98 , $p$ = .428, $\eta^2_p$ = .02.

In summary, the behavioural reaction time data from the online tasks indicated that the low worry group had a benign interpretation bias, evidenced by faster reaction time and higher endorsement rate to benign compared to negative trials, which was absent in the high worry group. The N400 amplitude data also indicated that individuals showed different interpretation biases between groups in the LDT, evidenced by greater N400 for negative than benign trials in low worry group and the high worry group, did not show any bias. However, the SWAT did not show the group differences on the EEG, failing to replicate the findings in the LDT.

## Discussion

This is the first study to explore interpretation bias in high and low worriers at different stages of information processing, from the early online interpretations reflected by the neurophysiological N400 index and behavioural measures, through to reflective offline interpretations. Findings in both offline measures and the online LDT were broadly consistent with our predictions: low worriers had a benign interpretation bias, whereas high worriers showed no interpretation bias. The only inconsistent finding was in the SWAT, in which only reaction time but not ERP results were consistent with our predictions. The current findings demonstrate that high and low worriers differ in their interpretative styles consistently across online and offline processing stages. High worriers had no interpretation bias, while low worriers had a benign interpretation bias.

The two offline measures of interpretation (scenario task and recognition task) both showed that high worriers did not have a preferential bias towards any type of interpretations, while low worriers had a clear benign interpretation bias. Our findings were consistent with a former study (Eysenck et al., 1991) that also used the recognition task with separate benign and

negative interpretation indices. On the other hand, previous studies that involved relative judgments instead of separate indices could not distinguish benign and negative interpretations and were interpreted as showing that the anxious group had more negative interpretation bias than non-anxious group (e.g., Anderson et al., 2012). The different conclusions between previous studies are likely due to paradigms not providing separate valenced indices rather than that the interpretation tendencies between participants in these studies were necessarily different. This indicates the clear benefits of paradigms that provide separate and independent indices for benign and negative interpretations.

The LDT online measure of interpretation bias showed that low worriers reacted faster to benign than to negative trials. This effect was less evident in high worriers, indicating they lack the benign interpretation bias that low worriers displayed. The findings from the LDT are consistent with other LDT studies (Hirsch & Mathews, 1997; 2000) and another study using a speeded grammatical decision task (Moser et al., 2012); they all showed that non-anxious groups had a benign interpretation bias that was lacking in the anxious populations. However, it should be noted that this conclusion was drawn from comparing benign and negative trials within each groups directly. The group by valence interaction in our LDT task did not reach statistical significance in contrast to earlier research (e.g., Hirsch & Mathew, 1997; 2000). This may be due to the compatibility of materials and populations, since the sensitivity of measuring interpretation in LDT depends on a good match between an individual's inferred concept and the provided interpretation (i.e., target word). The socially anxious population in the Hirsch & Mathew's (1997; 2000) studies are likely to have had more specific concerns relating to their performance in social situations; hence, the materials they used were narrowed down to social topics and pertinent to all participants, thus potentially driving the stronger effects. In our study, the worry topics were more diverse across worriers; therefore, the materials were based on a wide range of worry topics. It is likely that not all the materials were pertinent to all participants,

possibly diluting the effect. In relation to the SWAT, there were strong effects in endorsement rate (i.e., whether the word was related to the sentence) and reaction time (speed to make that decision) indices that showed high and low worriers had different interpretation biases. The effects were driven by low worriers showing a benign interpretation bias, whereas high worriers did not show any biases. Our endorsement rate results were consistent with the previous two studies (Amir et al., 2012; Beard & Amir, 2009) that used the original SWAT method, in which the words were presented preceding the sentences. Since the previous studies did not compare the reaction times to the negative and benign trials within a group, we could not compare the results between the current and former studies directly. However, our findings on low and high worriers appeared to be broadly consistent with studies on socially anxious populations. As with the offline measures, the approach we used is particularly informative as it allows comparison of negative and benign interpretation biases within a group.

Using the temporal precision available to ERP methods, and the N400 component to assess cognitive processing prior to the behavioural response, we found high and low worriers had different interpretation styles in the LDT, with a benign interpretation bias indicated in low worriers and no bias found in high worriers. This result was consistent with Moser et al. (2012), who used a similar task that required participants to make grammatical decisions. However, the results in the SWAT were inconsistent with the LDT in that there were no biases within the groups. This inconsistency in the two tasks may result from the different integrative levels of the sentences and the words in these tasks. In the LDT, the target words are the final words of the sentences. This design allows participants to read the sentences and the words as a whole and the sentences are designed to remain incomplete until the ambiguity is resolved. However, in the SWAT, the target words are not part of the sentences, they can be seen as words following sentences rather than complete information. Therefore, target words may be hard to integrate into the sentences, lessen the integrative levels, and yield similar N400 effects across groups.

The other possible explanation for the inconsistency is the requirements of the tasks. Previous research has shown that the basic task requirements (e.g. lexical decision, simple upper or lower case judgment) may affect N400 (Chwilla, Brown, & Hagoort, 1995; Silva-Pereyra et al., 1999). However, to our knowledge, there is no study comparing the N400 effects between lexical decision and relatedness judgement. The LDT requires participants to decide whether or not the final word is a real word, which is a simple linguistic judgment, and is not relevant to how participants interpret the sentences. On the contrary, the SWAT requires participants to make a relatedness judgment, which may require more mental resources and initiate an initial process to determine the match between the sentence and the word. Holding the goal of matching information in mind may interfere with the N400 effect, since N400 reflects a spontaneous process of the ease of information integration rather than an active intention to search or retrieve associated information. On the other hand, having the intention to search for possible interpretations is essential for the reaction time index when measuring interpretation in the SWAT, where the faster speed represents a closer match between target word and an individual's interpretation. This may explain the discrepancy between reaction time and N400 results in the SWAT.

Given that the Moser et al. (2012) study and our study did find evidence of interpretation bias by comparing the N400 amplitudes of different interpretations, N400 has the potential to be a sensitive tool for assessing early online interpretations in other clinical or non-clinical populations. However, factors that may affect the sensitivity of N400 should be explored in a future study. The discrepancy of the N400 results in two measures in the current study indicates that some paradigms may be more appropriate for use of N400 than others. One potential critical aspect involves the target words being integrated into the text to enhance the utility of using N400 as an interpretation index. Other factors that may influence N400 (e.g., task requirement) should be clarified in future studies, which may lead to novel ERP paradigms

being developed and being of use in exploring early interpretation bias in other populations.

One limitation of our study is that we did not control the lexico-semantic characteristics of target words, such as word length, word frequency, word concreteness or abstractness, which are factors that may affect the reaction time and N400 results. This is because the nature of material requirements was that sentences could be interpreted in benign or negative ways and that the target word was in keeping with one of these interpretations. This greatly limited options when selecting words and thus rendered matching of benign and negative targets on lexico-semantic characteristics very difficult. Indeed, the frequency of benign words was higher than the frequency of negative words in our online measures. Higher frequency words are generally processed faster than lower frequency words. Thus, based on objective word frequency, negative words should have had slower latencies. However, this was not the case in high worriers who had shorter latencies (although non-significantly so) to negative compared to benign word trials in the SWAT. Thus, the absence of an advantage for benign (higher frequency) words may partially supports a negative bias in high worriers, though the potential negative bias pattern in high worriers was not shown in the LDT. However, a negative bias was not evident in the LDT nor in the other two offline interpretation measures in the current study, and a lack of bias finding is consistent with the previous anxiety related interpretation bias literature (Eysenck et al. 1991; Hirsch & Mathews, 1997; 2000). Furthermore, as the study focused on the difference in interpretation bias between high and low worriers when they were exposed to the same materials, the conclusions were based on the differences between groups. Future studies will need to explore the potential frequency effect further, using word conditions matched for frequency and other lexico-semantic characteristics. The other limitation is that there were four non-native English speakers in the sample. Fluent non-native speakers may not show the same N400 effect as native speakers (Martin et al., 2013). However, we did not find any differences in the N400 results when excluding these four participants from both the LDT

and SWAT analyses. Another limitation is that there were six left-handed low worry participants included in this study, which may affect the topography of the N400. However, the Group by Electrodes Sites interactions of the LDT and SWAT showed no differences between groups in any electrode sites (*p*s > .324), despite the low worry group including more left-handed participants than the high worry group. Future studies should note this potential issue and include only right-handed participants.

The overall findings in the current study reveal an interesting phenomenon that high worriers had no interpretation bias, while low worriers showed a consistent benign bias across different interpretation processing stages. Interestingly, the N400 results showed that this phenomenon can be found at a neural level, indicating that low worriers generate benign interpretation in the early stage of information processing, and this benign interpretation is likely to remain active for prolonged periods and even accessible when individuals reflect on that information. The consistent pattern in both groups indicates that the interpretative tendency is evident when individuals first encounter ambiguity, and is sustained in the short term. The benign interpretation bias may serve as a protective factor against worry, since it is the major and consistent difference between high and low worriers across different stages of the interpretation processing. In addition, previous studies (Hayes et al., 2010; Hirsch et al., 2009) have shown that high worriers who were given consistently benign interpretations had fewer negative intrusions on a worry measure compared to controls after a single-session CBM-I program. A previous study also found that participants with GAD had increased offline benign interpretation bias and lower worry levels after a multi-session CBM-I program (Hirsch et al., 2018). Therefore, facilitating a benign bias may be helpful for reducing worry and changing the chain of interpretation processes. According to our results, it is likely that there is no need for worriers to have a negative bias at the beginning to benefit from CBM-I, which is different from previously held views in the CBM-I literature. However, no study to date has measured

ERPs to investigate if CBM-I also changes early online interpretations via single-session or multi-session CBM-I. It will be important to know whether enhancing benign offline interpretation by CBM-I could promote earlier online benign interpretations, and how many sessions are needed to change the early online interpretations, which may facilitate change in the interpretation process. It will also be interesting to know whether changes in all information processing stages is essential for reducing worry. Future studies could address these questions and explore the most efficient methods to alter interpretation bias and ultimately reduce symptoms of worry and anxiety.

An important finding in our study is that high worriers did not show any interpretation bias across different stages of information processing. One of two ways to explain this result is that they generate negative and benign interpretations equally; the other is that they did not generate any interpretation when encountering ambiguity. It would be interesting to explore which hypothesis is correct. If worriers generate both negative and benign interpretations, then it indicates that although the negative interpretations do not dominate the interpretation process, they may be sufficient to make high worriers unable to control their worry. On the contrary, if worriers do not generate any interpretations, then the lack of benign interpretation bias may facilitate continued worry. The current study investigated the relationship between worry and interpretations generated at the online and offline stages of interpretation processing. However, it is more likely that worry is maintained by the combined effects of a number of cognitive biases and cognitive processes (Hirsch, Clark, & Mathews, 2006; Hirsch et al., 2012). Future research could use online interpretation tasks with an ERP measure and other cognitive measures (e.g., memory, attention, attentional control) to explore how cognitive processes interplay with interpretation processing and contribute to worry. With the help of an appropriate interpretation paradigm for measuring the N400, the precise cognitive characteristics underlying worry and psychological disorders could be better understood by exploring how the

early interpretation tendency interacts with other closely associated cognitive processes.

In sum, our study showed that high worriers have no specific interpretation bias, while low worriers have a benign interpretation bias. These tendencies were found from the early online stage of interpretation represented by N400 to the later offline stage. Our findings have important implications, suggesting that the benign interpretation bias can be a protective factor against worry; the early online benign interpretation bias may remain active through to later offline interpretation processes. Future examination of methods to reduce worry perseveration should consider targeting and enhancing the early benign interpretation bias to examine how it affects interpretation at later stages in processing. Our study also implies that with more studies exploring factors that influence the utility of N400 in presenting interpretations, N400 can be a sensitive tool for understanding early online interpretation and the interactions between online interpretations and other cognitive processes.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations of conflict of interest: none

**Acknowledgments**

We are grateful to Suraya Dunsford and Calum Gordon for their help with recruitment, data collection and data checking, and to the participants for taking part in this research. YF receives scholarship from the government of Taiwan for her PhD study. CK receives salary support from an MQ: Transforming Mental Health PsyIMPACT grant. The authors declare no conflict of interest.

## References

American Electroencephalographic Society. (1991). *Guidelines for standard electrode position nomenclature*. Journal of Clinical Neurophysiology, *8*(2), 200-202.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders, fifth edition*. Arlington, VA: American Psychiatric Publishing.

Amir, N., Foa, E. B., & Coles, M. E. (1998). Negative interpretation bias in social phobia. *Behaviour Research and Therapy*, *36*, 945–957.

Amir, N., Prouvost, C., & Kuckertz, J. M. (2012). Lack of a benign interpretation bias in social anxiety disorder. *Cognitive Behaviour Therapy*, *41*(2), 119-129.

Anderson, K. G., Dugas, M. J., Koerner, N., Radomsky, A. S., Savard, P., & Turcotte, J. (2012). Interpretive style and intolerance of uncertainty in individuals with anxiety disorders: A focus on generalized anxiety disorder. *Journal of anxiety disorders*, *26*(8), 823-832.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12-28.

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1992). The speed of comprehension test: version A. *Speed and capacity of language processing test.* Suffolk: Thames Valley Test Company.

Barlow, D. H. (2002). *Anxiety and its disorders: The nature and treatment of anxiety and panic*. New York, NY: Guilford press.

Beard, C., & Amir, N. (2009). Interpretation in social anxiety: When meaning precedes ambiguity. *Cognitive therapy and research*, *33*(4), 406-415.

Bisson, M. S., & Sears, C. R. (2007). The effect of depressed mood on the interpretation of ambiguity, with and without negative mood induction. *Cognition and Emotion*, *21*(3), 614-645.

Brown, T. A., Antony, M. M., & Barlow, D. H. (1992). Psychometric properties of the Penn

State Worry Questionnaire in a clinical anxiety disorders sample. *Behaviour research and therapy*, *30*(1), 33-37.

Butler, G., & Mathews, A. (1983). Cognitive processes in anxiety. *Advances in behaviour research and therapy*, *5*(1), 51-62.

Chung, G., Tucker, D. M., West, P., Potts, G. F., Liotti, M., Luu, P., & Hartry, A. L. (1996). Emotional expectancy: Brain electrical activity associated with an emotional bias in interpreting life events. *Psychophysiology*, *33*(3), 218-233.

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, *32*(3), 274-285.

Davey, G. C. (1993). A comparison of three worry questionnaires. *Behaviour research and therapy*, *31*(1), 51-56.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

Eysenck, M. W., Mogg, K., May, J., Richards, A., & Mathews, A. (1991). Bias in interpretation of ambiguous sentences related to threat in anxiety. *Journal of abnormal psychology*, *100*(2), 144-150.

Grol, M., Schwenzfeier, A. K., Stricker, J., Booth, C., Temple-McCune, A., Derakshan, N., ... & Fox, E. (2018). The worrying mind in control: An investigation of adaptive working memory training and cognitive bias modification in worry-prone individuals. *Behaviour research and therapy*, *103*, 1-11.

Hayes, S., Hirsch, C. R., Krebs, G., & Mathews, A. (2010). The effects of modifying interpretation bias on worry in generalized anxiety disorder. *Behaviour research and therapy*, *48*(3), 171-178.

Hayes, S., Hirsch, C., & Mathews, A. (2008). Restriction of working memory capacity during worry. *Journal of abnormal psychology*, *117*(3), 712-717.

Hirsch, C. R., & Mathews, A. (1997). Interpretative inferences when reading about emotional events. *Behaviour Research and Therapy, 35*(12), 1123-1132.

Hirsch, C. R., & Mathews, A. (2000). Impaired positive inferential bias in social phobia. *Journal of abnormal psychology*, *109*(4), 705-712.

Hirsch, C. R., & Mathews, A. (2012). A cognitive model of pathological worry. *Behaviour research and therapy*, *50*(10), 636-646.

Hirsch, C. R., Clark, D. M., & Mathews, A. (2006). Imagery and interpretations in social phobia: Support for the combined cognitive biases hypothesis. *Behavior Therapy*, *37*(3), 223-236.

Hirsch, C. R., Hayes, S., & Mathews, A. (2009). Looking on the bright side: accessing benign meanings reduces worry. *Journal of abnormal psychology*, *118*(1), 44-54.

Hirsch, C. R., Krahé, C., Whyte, J., Loizou. S., Bridge, L., Norton, S., & Mathews, A. (2018). Interpretation training to target repetitive negative thinking in Generalized Anxiety Disorder and Depression. *Journal of Consulting and Clinical Psychology 86*(12), 1017-1030.

Hirsch, C. R., Meeten, F., Krahé, C., & Reeder, C. (2016). Resolving ambiguity in emotional disorders: the nature and role of interpretation biases. *Annual review of clinical psychology*, *12*, 281-305.

Krahé, C., Mathews, A., Whyte, J., & Hirsch, C. R. (2016). Cognitive bias modification for interpretation with and without prior repetitive negative thinking to reduce worry and rumination in generalised anxiety disorder and depression: protocol for a multisession experimental study with an active control condition. *BMJ open*, *6*(12), e013404.

Krahé, C., Whyte, J., Bridge, L., Loizou, & Hirsch, C. R. (2019). Are different forms of repetitive negative thinking associated with interpretation bias in generalized anxiety disorder and depression? *Clinical Psychological Science*, 2167702619851808. doi:10.1177/2167702619851808.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, *16*(9), 606-613.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in cognitive sciences*, *4*(12), 463-470.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, 8:213.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical care*, *46*(3), 266-274.

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574-588.

Mathews, A., & Mackintosh, B. (2000). Induced emotional interpretation bias and anxiety. *Journal of Abnormal Psychology, 109*, 602–615.

Mathews, A., Richards, A., & Eysenck, M. (1989). Interpretation of homophones related to threat in anxiety states. *Journal of abnormal psychology*, *98*(1), 31-34.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State worry questionnaire. *Behaviour research and therapy*, *28*(6), 487-495.

Mogg, K., Baldwin, D. S., Brodrick, P., & Bradley, B. P. (2004). Effect of short-term SSRI treatment on cognitive bias in generalised anxiety disorder. *Psychopharmacology*, *176*(3-4), 466-470.

Molina, S., & Borkovec, T. D. (1994). The Penn State Worry Questionnaire: Psychometric properties and associated characteristics. In G. C. L. Davey & F. Tallis (Eds.), *Worrying: Perspectives on theory, assessment and treatment* (pp. 265–283). New York, NY: Wiley.

Monsell, S. (1991). The nature and locus of word frequency effects. In *Basic processes in reading: Visual word recognition* (pp. 148–197). Hillsdale, NJ: Erlbaum.

Moreno, E. M., & Rivera, I. C. (2014). Setbacks, pleasant surprises and the simply unexpected: brainwave responses in a language comprehension task. *Social cognitive and affective neuroscience*, *9*(7), 991-999.

Moreno, E. M., & Vázquez, C. (2011). Will the glass be half full or half empty? Brain potentials and emotional expectations. *Biological psychology*, *88*(1), 131-140.

Moser, J. S., Hajcak, G., Huppert, J. D., Foa, E. B., & Simons, R. F. (2008). Interpretation bias in social anxiety as detected by event-related brain potentials. *Emotion*, *8*(5), 693-700.

Moser, J. S., Huppert, J. D., Foa, E. B., & Simons, R. F. (2012). Interpretation of ambiguous social scenarios in social phobia and depression: evidence from event-related brain potentials. *Biological psychology*, *89*(2), 387-397.

Murphy, R., Hirsch, C. R., Mathews, A., Smith, K., & Clark, D. M. (2007). Facilitating a benign interpretation bias in a high socially anxious population. *Behaviour research and therapy*, *45*(7), 1517-1529.

Oglesby, M. E., Raines, A. M., Short, N. A., Capron, D. W., & Schmidt, N. B. (2016). Interpretation bias for uncertain threat: A replication and extension. *Journal of behavior therapy and experimental psychiatry*, *51*, 35-42.

Ruscio, A. M. (2002). Delimiting the boundaries of generalized anxiety disorder: differentiating high worriers with and without GAD. *Journal of Anxiety Disorders*, *16*(4), 377-400.

Sears, C. R., Bisson, M. S., & Nielsen, K. E. (2011). Dysphoria and the immediate interpretation of ambiguity: Evidence for a negative interpretive bias in error rates but not response latencies. *Cognitive therapy and research*, *35*(5), 469-476.

Silva-Pereyra, J., Harmony, T., Villanueva, G., Fernández, T., Rodríguez, M., Galán, L., ... &

Reyes, A. (1999). N400 and lexical decisions: automatic or controlled processing?. *Clinical Neurophysiology*, *110*(5), 813-824.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine*, *166*(10), 1092-1097.

Stöber, J. (1998). Reliability and validity of two widely-used worry questionnaires: Self-report and self-peer convergence. *Personality and Individual differences*, *24*(6), 887-890.

Stopa, L., & Clark, D. M. (2000). Social phobia and interpretation of social events. *Behaviour Research and Therapy*, *38*, 273–283.

Suarez, L., & Bell-Dolan, D. (2001). The relationship of child worry to cognitive biases: Threat interpretation and likelihood of event occurrence. *Behavior Therapy*, *32*(3), 425-442.

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. *Oxford handbook of event-related potential components*, 397-440.

Tallis, F., Eysenck, M., & Mathews, A. (1992). A questionnaire for the measurement of nonpathological worry. *Personality and individual differences*, *13*(2), 161-168.

Voncken, M. J., Bögels, S. M., & de Vries, K. (2003). Interpretation and judgmental biases in social phobia. *Behaviour Research and Therapy*, *41*(12), 1481-1488.

**Figure captions**

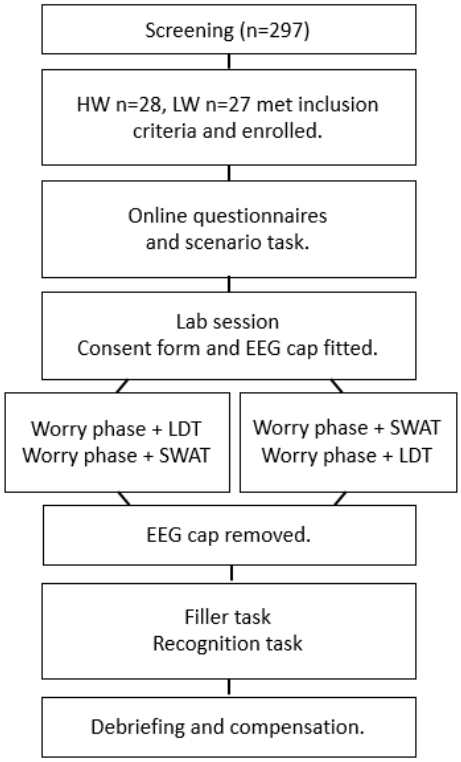
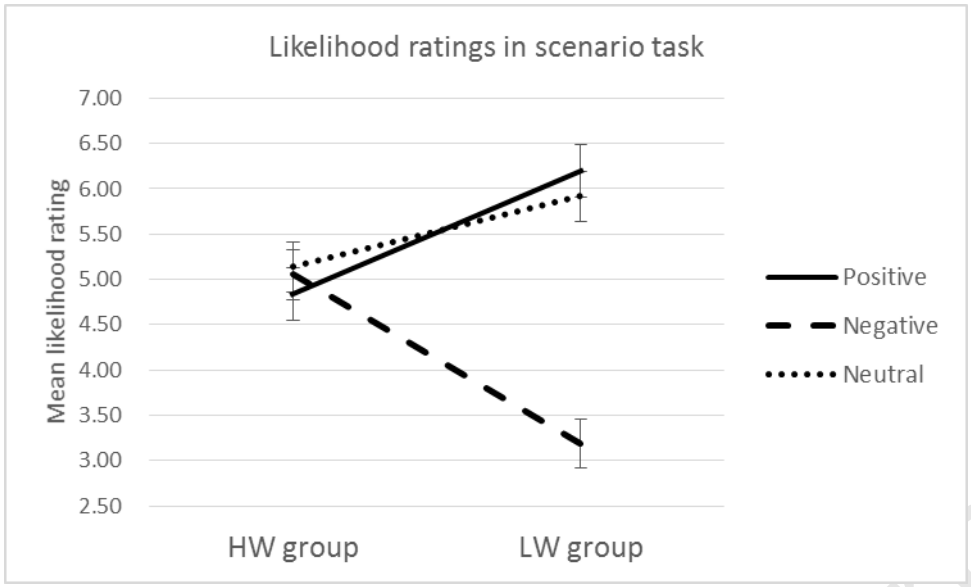


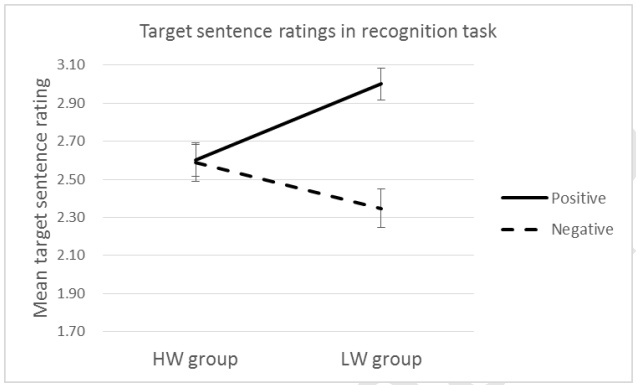

Figure 1.

Study flow chart.

(a)




(b)




(c)

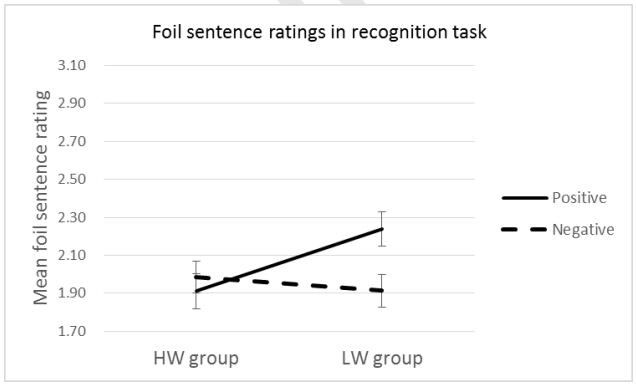
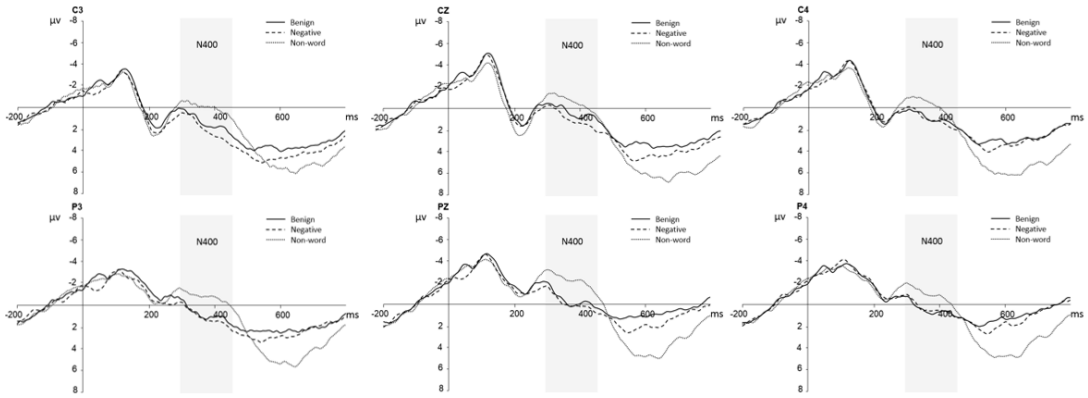



Figure 2.

Mean Likelihood ratings (1-9) in the scenario task (a) and mean similarity ratings (1-4) in the recognition task for target sentences (b) and foil sentences (c). Black lines are positive, dashed lines are negative, and the dotted line is neutral interpretation.

(a)




(b)

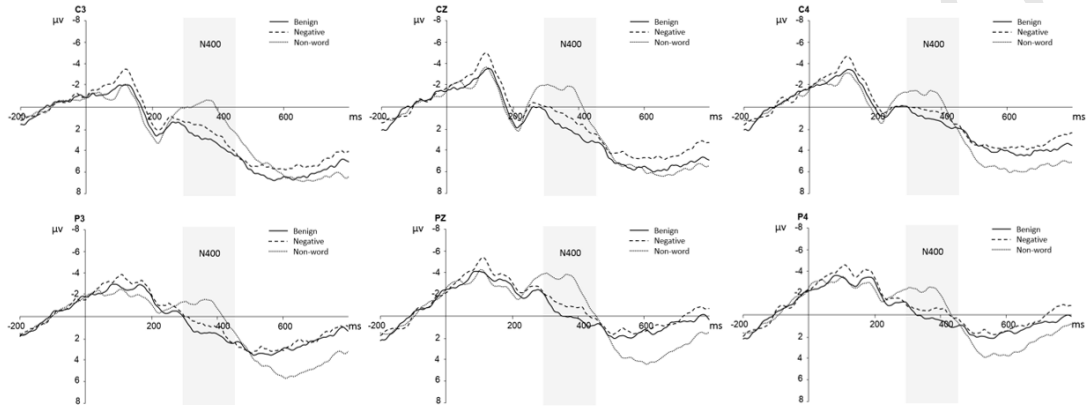
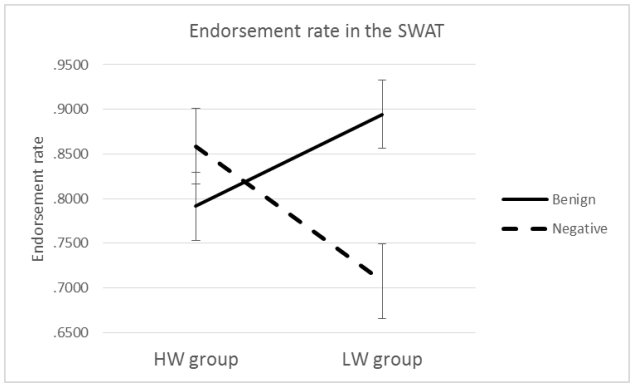


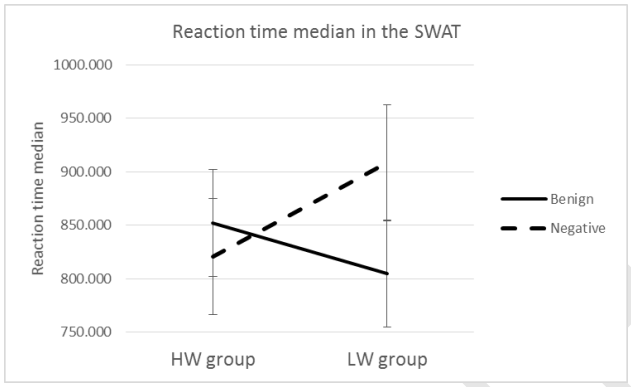

Figure 3.

N400 at C3, Cz, C4, P3, Pz, and P4 time-locked to benign (black), negative (dash) and non-word (dot) in the high worry group (a) and the low worry group (b) of the LDT. Negative voltage plotted up.

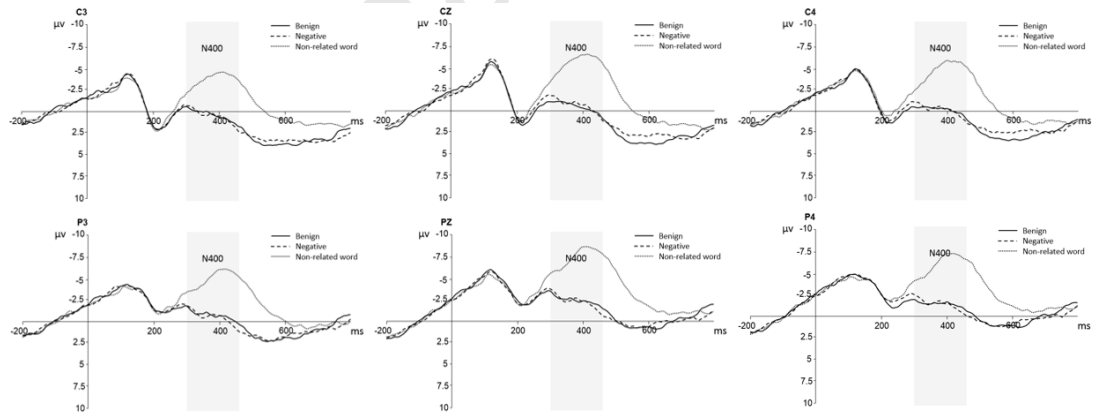
(a)




(b)




(c)




(d)

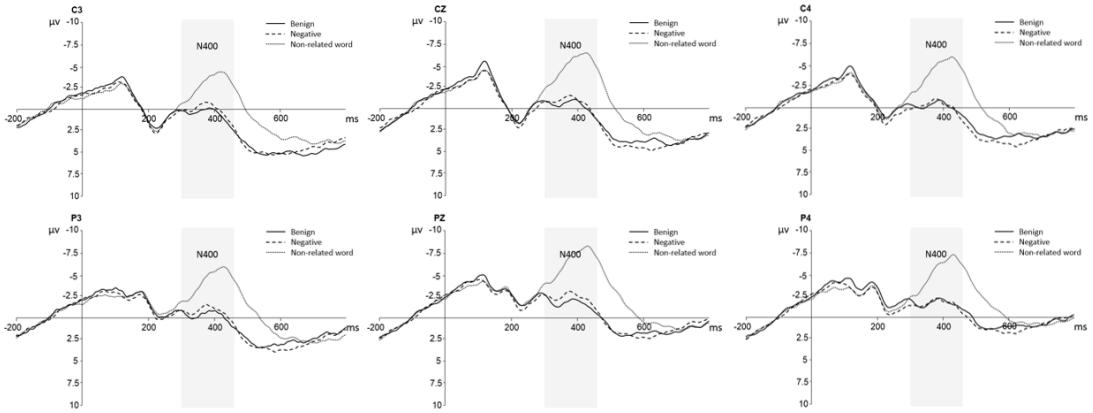



Figure 4.

Endorsement rate (a) and reaction time median (b) for benign (black) and negative (dash) words for two groups in the SWAT. N400 at C3, Cz, C4, P3, Pz, and P4 time-locked to benign (black), negative (dash) and non-related (dot) words in the high worry group (c) and the low worry group (d) of the SWAT. Negative voltage plotted up.

**Table**

Table. 1

Mean (standard deviation; range) and statistics for questionnaire scores.

| | High Worry group (n=28) | Low Worry group (n=27) | Test |
|---|---|---|---|
| Questionnaires | | | |
| PSWQ | 67.21 (6.18; 57-79) | 28.63 (6.36; 16-38) | $t$ (53) = 22.83, $p$ < .001 |
| GAD-7 | 10.96 (5.78; 1-21) | 1.00 (1.92; 0-9) | $t$ (53)= 8.51, $p$ < .001 |
| PHQ-9 | 10.82 (6.27; 2-26) | 1.11 (1.53; 0-6) | $t$ (53)= 7.82, $p$ < .001 |
| WDQ | 87.18 (20.25; 50-122) | 36.93 (7.87; 25-55) | $t$ (53)= 12.05, $p$ < .001 |

Notes. *PSWQ = Penn State Worry Questionnaire; GAD-7 = Generalized Anxiety Disorder 7-item scale; PHQ-9= Patient Health Questionnaire*; *WDQ = Worry Domains Questionnaires*.

Table 2.

Examples of material in four interpretation bias measures and their outcome measures.

| Ambiguous scenario or sentence in four tasks | Interpretations of the scenario or sentence | Interpretation bias measure |
|---|---|---|
| **Scenario task's scenario** | **Rating items** | |
| "*You have been visiting some friends in the centre of town, when you realise it is getting late. They offer you a lift but you set off on foot. Walking through a street that you don't know at all well, you can hear someone running up from behind.*" | Positive: "*In the unfamiliar street your friend runs up from behind to walk with you*".<br>Negative: "*In the unfamiliar street a mugger runs up from behind and threatens you*".<br>Neutral: "*In the unfamiliar street a woman is jogging*". | Ratings for the question: "*How likely would the outcome come to your mind?*" (1 very unlikely to 9 very likely). |
| **Recognition task's scenario** | **Recognition items** | |
| *"TITLE: Late return home*<br>*Your partner is working late this evening but now it is well past the time you were expecting them home. You are thinking about a crash you saw on the route your partner drives, when the phone rings. You pick it up and find out what had ha__en_d"* | Positive target interpretation: *"The phone rings and it is your partner telling you they are nearly home."*<br>Negative target interpretation: *"The phone rings and you are informed your partner is hurt in the accident."*<br>Positive foil: *"The phone rings and a friend invites you and your partner round for a meal."*<br>Negative foil: *"The phone rings and a friend tells you about gossip being spread about you."* | Ratings for the similarity of statements to the original scenario. (1 very different in meaning to 4 very similar in meaning). |
| **Lexical decision task sentences** | **Lexical decision items** | |
| "*As you give a speech, you see a person in the crowd smiling, which means that your speech is…*"<br><br>"*You are preparing for your annual work appraisal. You think that passing it will be…"*<br><br>*"You have been asked to look after* | Benign word in set A: "*funny*";<br>Negative word in set B: "*stupid*".<br><br>Benign word in set B: "*simple*",<br>Negative word in set A: "*hard*".<br><br>Non-word in set A and B: "*kejds*". | Reaction time for lexical decision. N400 amplitudes following words. |

| | | |
|---|---|---|
| *your niece because she is…"* | | |
| **Sentence word association task sentences** | **Association decision items** | |
| "*Your friends open your gift. You can tell how he feels by his face.*" | Benign word in set A: "*happy*"; Negative word in set B: "disappointed". | Endorsement rate and reaction time to endorse association items. N400 amplitudes following words. |
| *"There will be a role change in your office, and you have a meeting with HR."* | Benign word in set B: "*promoted*", Negative word in set A: "*fired*". | |
| *"You receive an unexpected grade on your test, which indicates your performance."* | Non-related word in set A and B: "*accommodate*". | |

Note: In the lexical decision task and the sentence word association task, the word that matches the benign interpretation of the sentence is called a benign word; the word that matched negative interpretation of the sentence is called a negative word.

Table 3.

The descriptive statistics in four interpretation bias measurements.

| | | | Descriptive statistics – mean (SD) | |
|---|---|---|---|---|
| Task | Interpretation indices | | High Worry group | Low Worry group |
| Scenario task | Likelihood ratings | | (n=28) | (n=27) |
| | Positive | | 4.84 (1.54) | 6.20 (1.53) |
| | Negative | | 5.05 (1.56) | 3.19 (1.32) |
| | Neutral | | 5.14 (1.44) | 5.92 (1.41) |
| | | | | |
| Recognition task | Similarity ratings | | (n=28) | (n=26) |
| | Positive target | | 2.60 (0.42) | 3.00 (0.47) |
| | Positive foil | | 1.91 (0.42) | 2.24 (0.53) |
| | Negative target | | 2.59 (0.48) | 2.35 (0.59) |
| | Negative foil | | 1.99 (0.45) | 1.91 (0.44) |
| | | | | |
| Lexical decision | RT median (ms) | | (n=25) | (n=25) |

| | | | | |
|---|---|---|---|---|
| task | Benign | | 650.64 (176.46) | 620.44 (110.04) |
| | Negative | | 679.46 (207.30) | 680.64 (152.11) |
| | | | | |
| | N400 amplitude (μV) (6 electrodes) | | (n=23) | (n=22) |
| | Benign | | 0.63 (0.74) | 1.43 (0.76) |
| | Negative | | 0.94 (0.73) | 0.63 (0.74) |
| | Non-word | | -0.75(0.79) | -1.09(0.81) |
| | | | | |
| Sentence word | Endorsement rate | | (n=26) | (n=27) |
| association task | Benign | | 79.13% (25.71%) | 89.44% (9.91%) |
| | Negative | | 85.87% (15.68%) | 70.74% (26.14%) |
| | | | | |
| | RT median (ms) | | (n=26) | (n=27) |
| | Benign | | 852.39 (294.08) | 805.07 (212.00) |
| | Negative | | 821.12 (268.50) | 908.22 (284.71) |
| | | | | |
| | N400 amplitude (μV) (6 electrodes) | | (n=22) | (n=23) |
| | Benign | | -0.82 (.67) | -0.61 (0.66) |
| | Negative | | -0.81 (0.73) | -0.95 (0.71) |
| | Non-related | | -5.35 (0.64) | -4.56 (0.63) |

Notes. Due to the nature of the SWAT. The numbers of endorsed trials varied between participants. The average was 81% (SD=16%), range from 55% to 100%.