**Cognitive correlates of mental health in adolescence:**

**A network analysis approach**

Sam Parsons, Annabel Songco, Charlotte Booth, and Elaine Fox

Department of Experimental Psychology

University of Oxford

**Authors note**

This work was supported by the European Research Council (ERC) under the European Union’s Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement no: [324176].

We would like to thank Jonas Haslbeck for his input on estimating moderated network models. We would also like to thank the other instructors at the Psychological Networks Amsterdam Winter School (2019) for their feedback on these analyses, and for suggesting a moderated network approach.

Correspondence concerning this article should be addressed to Sam Parsons, Department of Experimental Psychology, University of Oxford, New Radcliffe House, Radcliffe Observatory Quarter, Oxford, OX2 6AE. Email: [sam.parsons@psy.ox.ac.uk](mailto:sam.parsons@psy.ox.ac.uk)

The data and code can be found online: []

Cognitive correlates of mental health in adolescence: A network analysis approach

# Abstract

tbd

**Key words:** combined cognitive bias hypothesis, network analysis, adolescent, positive mental health

# Introduction

Automatic tendencies to selectively process negative, relative to benign or positive material, have been associated with anxiety and depression (for reviews, see Cisler & Koster, 2010; Gotlib & Joormann, 2010; Mathews & MacLeod, 1994, 2005). These biases have been explored in attention, interpretation of ambiguity, as well as in memory. Unfortunately, studies examining selective processing biases in relation to emotional vulnerability have tended to examine a single process in isolation, with few studies examining more than one bias in a single study (Everaert, Koster, & Derakshan, 2012; Hirsch, Clark, & Mathews, 2006). The combined cognitive bias hypothesis (CCBH: Hirsch et al., 2006), , proposes that cognitive biases do not work in isolation to influence emotional vulnerability, but rather, they influence each other and interact to influence other variables, including emotional vulnerability. For example, increased *attention* towards negative stimuli would be expected to influence how negatively information is *interpreted*, which would then influence *memory* for that stimulus. Such a series of causally related negative biases would be expected to propagate emotional vulnerability. Thus, greater influence of one bias on another may further help to perpetrate a negative cycle of maladaptive cognitive processing related to emotional dysfunction.

The CCBH has been tested with tasks designed to capture the direct effect of one bias on another. One study (Everaert, Duyck, & Koster, 2014) used an eye-tracking modification of a scrambled sentences task (Wenzlaff & Bates, 1998). Gaze fixation time on negative words was used to index attention bias, the ratio of negative to total unscrambled sentences was used to index interpretation bias, and a free recall task was used to index memory bias. Two path models were tested, the first omitted relationships between each of the biases, and the second included paths between the biases. The second model better fit the data, supporting the hypothesis that attention bias influenced interpretation bias, which in turn influenced memory bias. A further study in adults investigated the functional relationships among cognitive biases in a subclinical depressed sample and found that while attentional bias was not directly associated with memory bias, there was an indirect association via interpretation bias (Everaert, Tierens, Uzieblo, & Koster, 2013). Thus, the results supported the CCBH as cognitive biases across different domains are unlikely to act in isolation. A further study, conducted in adults, concluded that memory bias is likely to be more effectively modified by targeting emotional processing in another domain, such as interpretation bias (Vrijsen, Becker, Rinc, van Oostrom, Specken, Whitmer, & Gotlib, 2014; also see Hertel & Mathews, 2011).

To our knowledge only two studies have examined the combined cognitive bias hypothesis in adolescent samples (Klein, de Voogd, Wiers, & Salemink, 2017; Orchard & Reynolds, 2018). Klein et al. (2017) used three cognitive bias measures; an emotional visual search task, a dot-probe task, and the interpretation recognition task for children. The results indicate that each cognitive bias predicted unique variance in anxiety and depression, separately, supporting the CCBH proposition that cognitive biases in different domains contribute separately to emotional vulnerability. Similarly, a study by Orchard and Reynolds (2018) in adolescents showed that the combination of cognitive biases, interpretation bias and negative self-evaluation bias, predicted depression severity more strongly than individual biases. Both studies suggest that further exploration of adolescent mental health from a CCBH perspective is likely to be highly informative.

In addition, most relevant to the current study, the CCBH suggests we would expect to observe differences in relationships biases between clinical and non-clinical groups. The implication is that the interrelationships and interactions among biases themselves contribute to psychopathology. In one study, the factor structure of cognitive bias measures in attention, inhibition, imagery, and memory were found to differ between a formerly depressed, and non-clinical sample (Vrijsen, van Oostrom, Isaac, Becker, & Speckens, 2014). Coherence between attention and memory bias was found in the non-clinical sample, but not in the formerly depressed sample. This suggests that relationships among cognitive biases may differ between vulnerable and less-vulnerable groups, in support of the CCBH. In the present study, we examine these interrelationships in a non-clinical sample of adolescents.

### Psychological Network approaches

The CCBH has a striking overlap with a psychological network approach to emotional disorders. A network perspective on psychopathology views emotional disorders, such as anxiety and depression, as a system of interacting symptoms (Fried, 2017; Fried et al., 2017). As such, rather than individual symptoms acting alone to influence a disorder, the interrelations between them also play a key role. Borsboom and colleagues have been the driving force behind initiating network analyses in clinical psychology (Borsboom, Cramer, Schmittmann, Epskamp, & Waldorp, 2011; Schmittmann et al., 2013), and this has resulted in the application of network analysis approaches to psychopathology (e.g. Bernstein, Heeren, & McNally, 2017; Borsboom & Cramer, 2013; Heeren & McNally, 2016; McNally et al., 2015). A common aim of network approaches is to identify plausible, and potentially causal, connections amongst individual symptoms of a disorder (e.g. McNally et al., 2015).

A network theory of mental disorders has been proposed (Borsboom, 2017), including several core principles particularly relevant to the CCBH. These principles are comparable to the, albeit less formal, core elements of the CCBH; namely that we would expect different biases to interact with one another to influence emotional vulnerability and that biases may reinforce one another reciprocally to influence emotional vulnerability. The similarities are such that theoretically applying the principles of the network approach may prove fruitful and provide a more formal and systematic approach to investigate the CCBH. We propose that cognitive biases directly and indirectly influence one another (Borsboom’s *direct causal connections* principle), and that these interactions among different biases are likely to influence emotional vulnerability as well as emotional wellbeing (Borsboom’s *complexity* principle).

Network analysis enables the quantification and visualisation of the multivariate dependencies that exist in the dataset. In psychological network, *nodes* that represent observed psychological variables (e.g., psychometric tests or indices of cognitive bias) are connected by *edges,* which represent the observed statistical relationship between them, e.g. regularised partial correlations. The edge colour is a useful indication of the nature of the association; here positive associations are presented in blue and negative associations in red. The *edge weight* is used to indicate the strength of a relationship; stronger relationships are represented with thicker edges, whereas weaker relationships are denoted with thinner less saturated edges. In this paper we use a moderated network analysis to investigate mental health related changes in the structure of cognitive bias networks.

Network analyses have been used previously to investigate laboratory measures of cognition and behaviour, going beyond self-report measures (Bernstein et al., 2017; Heeren & McNally, 2016). Heeren and McNally investigated the interplay between social anxiety symptoms, attentional bias, and attentional control. Their analysis indicated that the orienting of attention was strongly linked to self-reported fear of social situations, which in turn was strongly related to avoidance of those situations. This has a potentially important clinical implication in that it suggests that interventions targeting attention orientation would positively influence other processes and propagate those benefits throughout the psychological network, resulting in therapeutic effects (McNally et al., 2015). Similarly, Bernstein et al. investigated components of executive control and rumination. The analysis suggested that self-criticism was central to the network with strong down-stream effects on negativity and brooding. Thus, reducing self-criticism may have a wide reaching beneficial effect on other components of rumination, and represents a potentially useful therapeutic target. Network analysis approaches provide an informative perspective on the interplay between cognitive processes and components of psychopathology and may help to inform the development of novel clinical interventions.

### Cognitive bias approaches to positive mental health and resilience

While information-processing approaches have been widely used to investigate the cognitive mechanisms of emotion dysfunction (for reviews, see Gotlib & Joormann, 2010; Lau & Waters, 2017; Mathews & MacLeod, 2005; Yiend, 2010) relatively little research has examined the role of selective information processing in positive mental health in adults (Carl, Soskin, Kerns, & Barlow, 2013; Parsons, Kruijt, & Fox, 2016) and even less in adolescents. Positive mental health and mental illness are considered to represent two distinct, albeit inversely correlated, continua (Keyes, 2002, 2005). Mental illness and low mental health have been found to have additive adverse effects on an individual’s functioning in life, including academic impairment and suicidal ideation (Keyes et al., 2012), as well as all-cause mortality (Keyes & Simoes, 2012). An implication of the dual continua model is that they may be characterised by distinct patterns of selective processing styles or biases, just as the ‘symptoms’ of mental health and mental illness differ from each other.

While some research has examined factors related to positive mental health within a cognitive-experimental framework, as with the CCBH, the majority of this research has been conducted in adults. Therefore, in this study we employed a psychological network analysis approach to investigate the CCBH in positive mental health in adolescents. We use data from the CogBIAS longitudinal study (Booth et al., 2017; Booth et al., 2019), one of the few studies to collect data for multiple cognitive biases in an adolescent sample (Klein et al. 2017; Orchard et al. 2018). We focussed on attention, interpretation, and memory bias, to limit the scope to the processes previously implicated in the CCBH (Hirsch et al., 2006). Our primary aim was to explore differences in the structure of cognitive bias networks between adolescents higher and lower in mental health. We therefore use a moderated network approach, which allowed us to examine the moderating effect of mental health on the relationships between other biases. To our knowledge, this is the first study to use network analyses to examine the role that connections in selective processing of emotional information plays in positive mental health in an adolescent sample.

# Methods

The data analysed and presented in this paper are drawn from the CogBIAS longitudinal study (Booth et al., 2017; Booth et al., 2019), which recruited approximately five hundred 12-14 year old secondary school adolescents in the UK. Adolescents completed a series of cognitive bias measures (including attention, interpretation, and memory) and two further waves of testing at the age of 14 and again at the age of 16. This study is the only one known to the authors to incorporate a longitudinal design, with a range of cognitive biases measured, at three time-points in an adolescent sample. The CogBIAS study presents an opportunity to examine the CCBH as it applies to adolescents, specifically with respect to the role these cognitive biases play in positive mental health. In this paper, we use data from wave 1 of the CogBIAS study.

### Participants

We first excluded all participants without complete data in all of the measures described below (except the dot-probe, see below), from the original sample of 504 students. This resulted in a final sample of 451 students (*M age* = 13, *SD* = .78, 248 female). In addition, for the first stage of our analysis we selected two groups of participants for the network analyses based on scores on the Mental Health Continuum (MHC). For this, we performed a tertile split to yield low and high Mental Health groups (low-MH and high-MH, respectively). The low-MH group consisted of 145 participants, scoring under 37 on the MHC and the high-MH group consisted of 150 participants scoring above 47 on the MHC. Ethical approval for this study was given by the National Research Ethics Service (NREC; REC reference: 14/SC/0128; IRAS project ID: 141833).

**Measures**

As we were interested in differences in network structure of cognitive biases in relation to high and low positive mental health we analyse only a subset of the measures included in the CogBIAS study. The combined cognitive bias hypothesis typically describes the relationship between attention, interpretation, and memory biases. We therefore analysed data only from tasks targeting these cognitive processes. We present a brief description of these measures below, however, a complete description of the sample, methods, and design used in the study can be found in the protocol paper (Booth et al., 2017).

**Mental health.** The Mental Health Continuum – Short form (MHC-SF) contains 14 items that index emotional, psychological, and social wellbeing, in order to create a composite measure of positive mental health. Participants are asked to rate how often they have experienced each of the items in the past month, on a 6-point Likert scale from "never" to "every day". The MHC-SF has shown high internal consistency and discriminant validity (Keyes, 2009; Lamers, Westerhof, Bohlmeijer, ten Klooster, & Keyes, 2011; in the current sample MacDonald’s Omega ω = .95, and Cronbach’s alpha α = .94, 95% CI [.93, .95]).

**Attention Bias.** An emotional face (angry, happy, and pain) dot-probe task was used to index attention bias (MacLeod et al., 1986) with stimuli from the STOIC faces database (Roy et al., 2007). As with other papers using data from the CogBIAS study, we have opted to omit the attention bias data from our analyses. The internal consistency of each of the attention bias indices (*n* = 448 following removal of three participants for < 70% accuracy) was below any acceptable threshold, in this sample; angry = -.07, 95% CI [-.20, .06]; happy = .12, 95% CI [.00, .24]; pain = -.04, 95% CI [-.17, .09]. These outcome measures are unsuitable for any analyses based on correlational measures and were omitted from any further analyses. For full details about the task, see Booth et al. (2017; 2019).

**Interpretation bias.** The Adolescent Interpretation and Belief Questionnaire (AIBQ; Miers, Blöte, Bögels, & Westenberg, 2008)contains ten hypothetical scenarios, five of which are socially oriented and five are non-socially oriented, that are intended to reflect events that are likely to be experienced by adolescents. Participants read the scenario and were presented with a question that addresses a point of ambiguity in the scenario. A positive, a neutral, and a negative interpretation of the scenario are presented and participants rate how likely that interpretation would pop into their mind on a 5-point Likert scale. Participants then choose which interpretation of the scenario they believe to be the most correct. Scenarios are presented in a pseudo-random order. Bias scores were computed by calculating the mean likelihood ratings for positive and negative interpretations of social and non-social situations separately, resulting in four bias indices – social positive; non-social positive; social negative; non-social negative. The reliabilities of each bias indices were as follows; social positive (α = .55, 95% CI [.48, .61], ω = .63); Social Negative (α = .79, 95% CI [.76, .82], ω = .81); Non-Social Positive (α = .43, 95% CI [.35, .52], ω = .49); Non-Social Negative (α = .58, 95% CI [.52, .64], ω = .64).

**Memory bias.** In the Self-Referential encoding task (SRET), participants were presented with 22 positive and 22 negative words in a random order (the word lists were drawn from Hammen & Zupan, 1984). Each word was presented for 200ms before a prompt "Describes me?" was presented on screen, after which participants responded ‘yes’ or ‘no’ using the ‘Y’ and ‘N’ keys on the computer keyboard. After all words had been presented, a short distraction task was administered consisting of three simple mathematics questions. Finally, in the incidental recall phase, participants were given three minutes to recall and type in as many words as they could remember. Positive and negative memory bias indices were calculated as the number of positive and negative words, respectively, that were endorsed (participants responded that the word described them) and subsequently recalled (Asarnow, Thompson, Joormann, & Gotlib, 2014). We are not aware of a suitable procedure for estimating the reliability of the memory bias indices due to the process of calculating the scores. Yet, the test-retest reliability (ICC3k) of the memory bias indices over three waves of testing was .68 and .72 for negative and positive memory bias, respectively, suggesting that the measure is somewhat reliable.

### Procedure

Participants were tested in groups ranging between 13 and 50 students in computer labs either in their own school, or at the University of Oxford. Testing consisted of two, one-hour sessions which were either back-to-back or on different days, depending on school and testing space availability. In each session, participants completed three tasks, in the same order, followed by a battery of questionnaires (see Booth et al, 2017, for further information on measures not analysed in this paper). Participants were asked to complete both sessions in exam conditions, i.e. not talking or looking their peers computer screens. At least one researcher was present throughout the testing sessions to answer any questions and ensure adequate testing conditions were maintained.

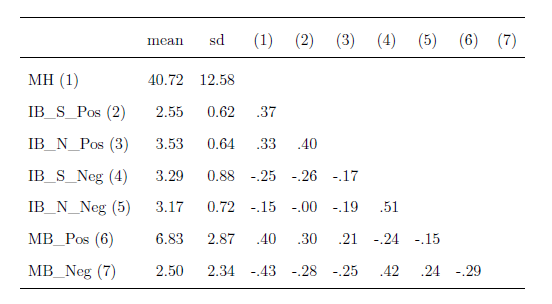
### Data analysis

In our preliminary analysis we compared the network structure of interpretation and memory biases for a high-MH and a low-MH group, following a tertile split of the data by mental health. These Gaussian graphical models formed the basis of the earlier version of this paper (in the next section we explain why a different approach is preferable). We computed a ‘graphical LASSO’ (glasso; Epskamp & Fried, 2016, also Friedman, Hastie, & Tibshirani, 2008) estimation procedure with EBIC model selection (Foygel & Drton, 2010). The glasso algorithm is implemented in the *glasso* package (Friedman, Hastie, & Tibshirani, 2014), and is called by the *bootnet* package (Epskamp et al., 2017), which we used for this paper. The glasso algorithm estimates a partial correlation network by directly penalising elements of the variance-covariance matrix. We set the tuning parameter gamma to 0.5 to generate a sparser network, due to the removal of potentially spurious associations. We then used the *NCT* function from the *NetworkComparisonTest* package (van Borkulo, 2016) to compare our high mental health and low mental health networks. The function tests for differences in the overall connectivity (as the sum of all edge weights in the network, or global strength) between networks.

For the core analysis, we conducted a moderated network analysis (Haslbeck, Borsboom, & Waldorp, 2019) using the *mgm* package, treating Mental Health as a moderating variable in the network. This improves on our first analysis in several ways. First, we were able to treat mental health as a continuous variable, rather than dichotomising our sample. As a result, were then able to use our full sample, retaining the third omitted from the previous analysis. Second, moderated networks examine moderation of individual edges, providing a more nuanced perspective on any moderating effects of mental health. This improves on the network comparison test, which only provides information about the difference in global connectivity between the networks. Haslbeck et al. demonstrated that a moderated network approach outperforms split-sample methods like network comparison test and fused graphical lasso models. We also obtained predictability indices for each node in the network. We then resampled the estimated network 500 times to obtain confidence intervals around the estimated network structure and moderation effects, thus providing vital information on the stability of the networks. The resampling procedure also allowed us to extract the proportion of non-zero edges and moderating effects. Finally, we computed and visualised three networks conditioned on mental health to illustrate changes in network structure at different levels of mental health.

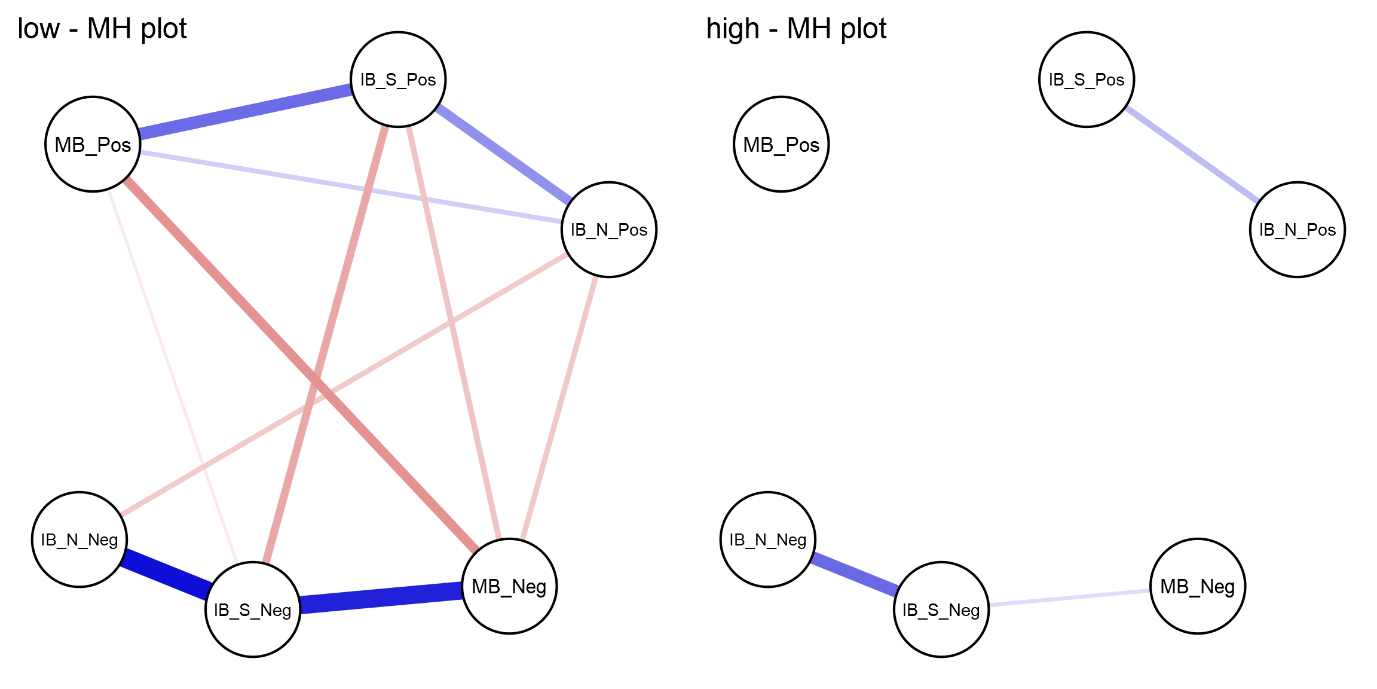
# Results

Table 1 presents the descriptive statistics and correlation matrix for all variables included in the network analysis. Correlation and covariance matrices for the full sample (S1-2), and the low and high MH subsample (S3-6) can be found in the supplemental materials.

Table 1. Descriptive statistics and correlation matrix for the full sample (n = 451). 

Note: MH = Positive mental health; IB\_S\_Pos = Social Positive Interpretation Bias; IB\_S\_Neg = Social Negative Interpretation Bias; IB\_N\_Pos = Non-Social Negative Interpretation Bias; IB\_N\_Neg = Non-Social; MB\_Pos = positive memory bias; MB\_Neg = negative memory bias.

**Comparing networks of cognitive biases in a low mental health group to a high mental health group.** First we estimated a graphical LASSO network (tuning parameter gamma was set to .5 to generate a sparser network) for the high and low mental health groups separately. Figure 1 presents a visualisation of both networks. In the low mental health network each node is connected to two or more other nodes; negative and positive biases are negatively associated; and, the strongest edges connect memory biases with social interpretation biases. In contrast, the high mental health network is substantially less interconnected compared to the low mental health network with only three retained edges compared to eleven. In the high mental health network, no negative relationships were retained between the positive and negative cognitive biases. Additionally, the edges retained in both networks appear weaker in the high mental health group. To formally compare the global strength of each network (the sum of edge strengths in the network) we used the *NCT* function from the *NetworkComparisonTest* package (van Borkulo, 2016). We ran 1000 iterations resampling from the networks. The low-MH network (global strength = 1.70) was more strongly connected overall than the high-MH network (global strength = .37), and this difference was statistically significant, *p* = .014.



*Figure 1.* Graphical LASSO Networks. The left and right panels present the graphical LASSO network from the low mental health group and the high mental health group, respectively. Each node represents a cognitive bias measure and each edge represents the (partial) correlation between the nodes it connects, after controlling for all other variables in the network. Larger correlations are represented by thicker edges. Blue edges indicate positive relationships, whereas red edges indicate negative relationships.

Note: IB\_S\_Pos = Social Positive Interpretation Bias; IB\_S\_Neg = Social Negative Interpretation Bias; IB\_N\_Pos = Non-Social Negative Interpretation Bias; IB\_N\_Neg = Non-Social; MB\_Pos = positive memory bias; MB\_Neg = negative memory bias.

**Moderated Network Analysis.** As described in the data analysis section, there are some limitations to the previous analysis. Splitting the sample and comparing the global strength provides some, but limited, information about the moderating effect of mental health on the interrelationships among cognitive biases. To address the question of moderation fully we analysed the full data set using a moderated network analysis approach (Haslebeck et al. 2018). We used the *mgm* package (Haslbeck, Borsboom, & Waldorp, 2018) to estimate the moderated network with positive mental health as the moderating variable. We used cross-validation to select the regularization parameter. Weak edges are shrunk to zero leading to a matrix of regularised coefficients representing conditional dependence relations. For the nodewise regressions *k* edge weights are obtained for each *k*-order interaction (e.g. 2 for pairwise interactions and 3 for the moderated effects). We used the OR-rule to combine these weights, which takes the mean of all *k* parameter estimates, as the default option and because the AND-rule may be too conservative for the 3-way interactions of interest (Haslbeck, Borsboom, & Waldorp, 2018). We then extracted predictability indices for each variable following (Haslbeck & Waldorp, 2018). Predictability refers to the proportion of variance explained by all other nodes in the network.

The resulting network of pairwise interactions is visualised in Figure 2 (for ease of interpretation we present the moderation effects separately in Figure 3). Mental health was connected most strongly to memory biases – i.e., a negative association with negative memory and positive association with positive memory. For interpretation biases, only edges connecting mental health to positive interpretation biases (social and non-social) were retained. Edges connecting mental health to negative interpretation biases were not retained. Non-social interpretation biases were not directly connected to memory biases (both positive and negative), but were connected via social interpretation biases. The explained variance (predictability metric) of mental health was 33.4%; the predictability of the cognitive biases ranged from 21.5% (positive memory bias) to 38.1% (negative social interpretation bias).

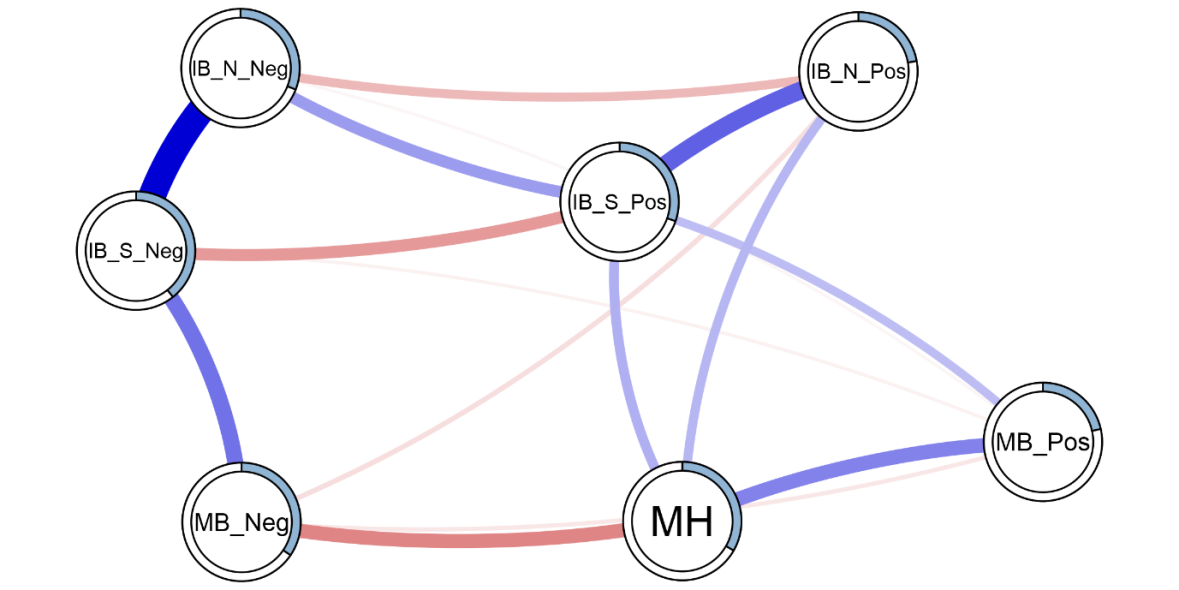


Figure 2. Estimated moderated network of positive mental health, and interpretation and memory cognitive biases. Blue edges represent positive associations, red edges represent negative associations; the width of the edge indicates the strength of this relationship. The shaded area of the pie surrounding each node represents the predictability of that variable, i.e., the variance explained by all other variables in the network. Note that this figure does not visualise the degree of moderation in the networks.

Note: MH = Positive mental health; IB\_S\_Pos = Social Positive Interpretation Bias; IB\_S\_Neg = Social Negative Interpretation Bias; IB\_N\_Pos = Non-Social Negative Interpretation Bias; IB\_N\_Neg = Non-Social; MB\_Pos = positive memory bias; MB\_Neg = negative memory bias.

To obtain information about the stability of the moderated network we resampled the moderated network 500 times. This allowed us to obtain confidence intervals surrounding each individual edge weight. It further allowed us to examine the strength of moderation for each edge and compute confidence intervals around the moderation estimate and the proportion of moderated edges. Figure 3 presents the mean weight and Confidence Intervals for individual edges. Additionally, Figure 3 presents the degree of moderation due to mental health on each edge as a result. The moderating effect of mental health on most edges follows a similar pattern the original analysis. Most edges were moderated towards zero with increased positive mental health. Thus, overall we would expect a sparser network at higher mental health, compared to low mental health. The main exception to this trend was the edge connecting social positive and non-social positive interpretation biases. This was the only (likely non-zero) edge to be moderated to be stronger with increases in mental health.

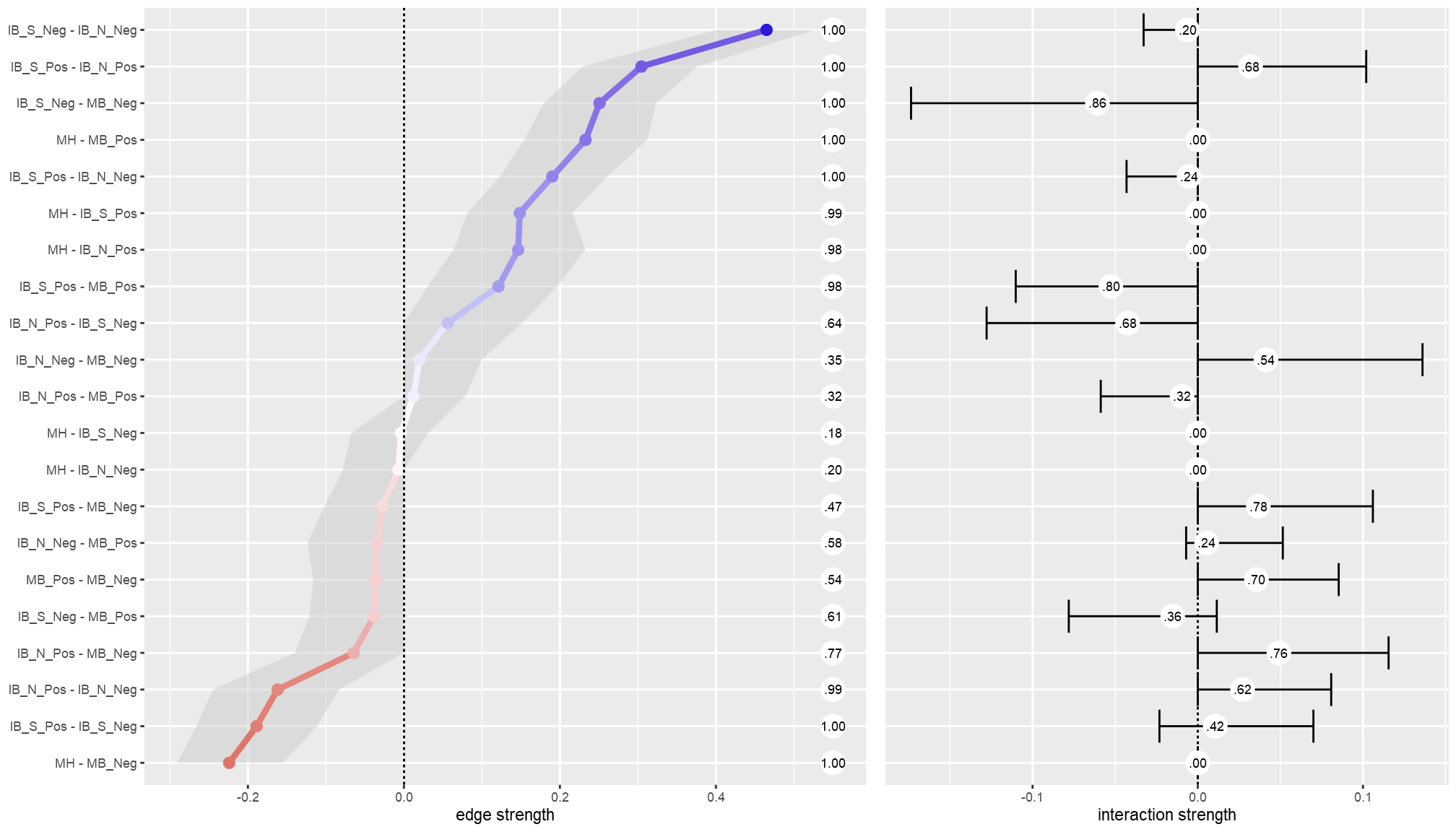


Figure 3. Edge strength and degree of moderation by mental health. The left panel presents the estimated edge strengths within the network (these correspond to the network visualization in figure 2) from 5000 resamples of the mgm network estimation procedure. The shaded area represents the 95% CI around the estimate. Numbers running down the centre of the figure represent the proportion of non-zero estimates for each edge. The right panel presents the estimated moderating effect of mental health on each edge; the ticks represent the 95% CI around the estimate. The circled numbers represent the number of non-zero moderation effects arising across the resamples.

Note: MH = Positive mental health. From the Adolescent Interpretation and Belief Questionnaire (AIBQ): IB\_S\_Pos = Positive interpretation bias in social scenarios; IB\_S\_Neg = Negative interpretation bias in social scenarios; IB\_N\_Pos = Negative interpretation bias in non-social scenarios; IB\_N\_Neg = Negative interpretation bias in non-social scenarios. From the self-referential encoding task (endorsed and recalled items): MB\_Pos = positive memory bias; MB\_Neg = negative memory bias.

To highlight the influence of mental health as the moderator of the networks, Figure 4 presents three networks comparing values of mental health. We used the *condition* function from the *mgm* package (Haslbeck & Waldorp, 2018) to condition the estimated moderated network (Figure 2) on the mental health moderator. We conditioned the network to -1 standard deviation from the mean, the mean, and +1 standard deviation from the mean on positive mental health (the left, centre, and right panels in figure 4, respectively).

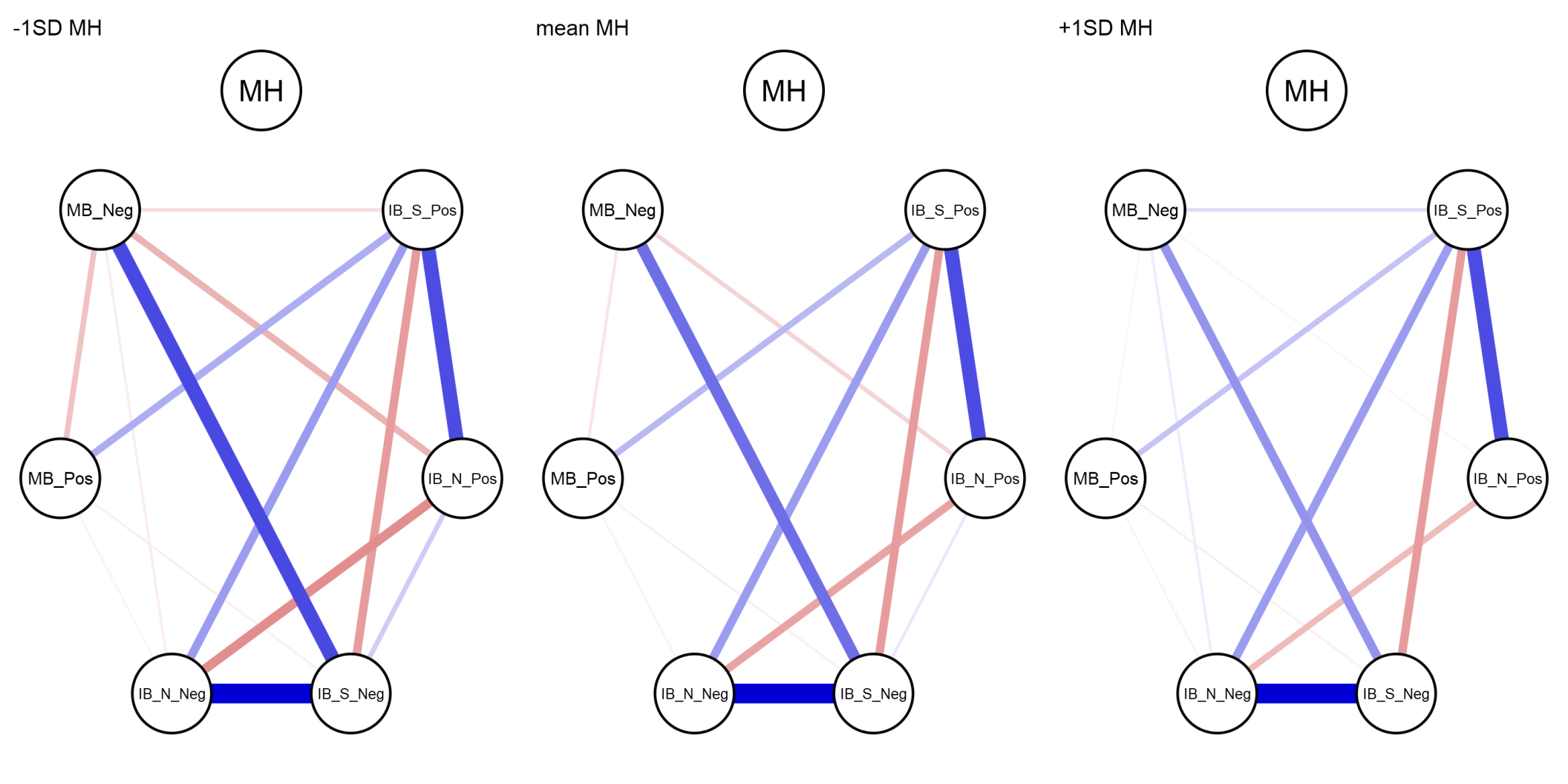


Figure 4. Networks of cognitive biases conditioned on mental health. Mental health is specified at -1SD (left) and +1SD (right) from the mean, and at the mean MH (centre). Conditioning sets mental health to the specified value, and therefore the edges connecting mental health are set to zero (as there is no variance in mental health).

Note: MH = Positive mental health. From the Adolescent Interpretation and Belief Questionnaire (AIBQ): IB\_S\_Pos = Positive interpretation bias in social scenarios; IB\_S\_Neg = Negative interpretation bias in social scenarios; IB\_N\_Pos = Negative interpretation bias in non-social scenarios; IB\_N\_Neg = Negative interpretation bias in non-social scenarios. From the self-referential encoding task (endorsed and recalled items): MB\_Pos = positive memory bias; MB\_Neg = negative memory bias.

# Discussion

The present study sought to investigate the Combined Cognitive Bias Hypothesis (CCBH; Hirsch et al., 2006; Everaert et al. 2012) in adolescents using a network approach. We analysed baseline data (at age 12-14 years) from the CogBIAS longitudinal study (Booth et al. 2017; 2019), including interpretation bias and memory bias measures. We excluded attention bias indices from the analyses due to low reliability (we discuss this in the limitations below). First, we split the sample into high and low mental health subsets. We estimated networks of biases for each group and compared the global strength of each network. The low mental health network showed greater connectivity than the high mental health group. We expanded on this analysis with a moderated network approach treating mental health as a moderating variable in the network of cognitive biases. The results suggest some, albeit small, moderation effects. The edge connecting negative social interpretation bias and negative memory bias was the most moderated edge, both in strength of moderation and in the proportion of resamples that included a non-zero moderation effect estimate. Mental health moderated each edge in 20% or more of the resamples (see Figure 3). Qualitatively, there was a trend for higher mental health to moderate edges towards zero, and therefore, a more weakly connected network. We interpret these moderation effects with some caution. Although some edges showed a relatively consistent pattern of moderation by mental health, no edge uniformly showed this moderation across resamples. Moderation effects are typically small (e.g. Haslbeck, Borsboom, & Waldorp 2018), which may explain the lack of stability of some moderation effects. Increased stability of the moderated network models would require larger samples.

**Implications for the Combined Cognitive Bias Hypothesis**

The CCBH includes questions of association, causality, and predictive magnitude, concerning how multiple cognitive biases may contribute to mental health (Everaert et al. 2014). We discuss our results in light of each CCBH question.

Our analyses address the CCBH *association* question in the broadest sense by showing interconnectedness amongst interpretation and memory biases. Memory biases were most strongly connected to mental health, and there were no direct edges connecting mental health and negative interpretation biases. Further, our moderation network model suggests that mental health was related to the network structure itself, in addition to individual biases. Our initial analysis suggested moderation of the network; our low mental health network was significantly more strongly connected than the high mental health network (Figure 1). We improved on this with a moderated network model. The moderated network indicates some moderation of the network structure by mental health (Figure 3). This supports the CCBH hypothesis that the interrelationships amongst cognitive biases may influence mental health, in addition to any direct relationships between cognitive biases and mental health.

To address CCBH *causality* questions, we need experimental tasks that integrate multiple cognitive bias processes (e.g. Everaert et al. 2014), or extensive longitudinal research designs. We cannot make inferences from our networks on the causal relations between biases, nor causal relations between the network structure of biases and mental health. Yet, network analyses are extremely useful for hypothesis generation. One way we can propose causal hypotheses is to examine edges that were not retained in the network. For instance, mental health was not directly connected to negative interpretation biases, when controlling for all other variables in the network. This suggests negative interpretation biases influence mental health via other biases in the network, e.g. negative memory bias. This follows from the CCBH and suggests a causal chain from negative interpretation bias for social scenarios, via negative memory bias, to influence mental health. Others have argued that it is difficult to modify memory bias directly, and training interpretation bias, with the intent to modify memory biases, may be more effective (Vrijsen et al. 2014). Related to this, our network does not include edges connecting non-social interpretation biases and memory biases. A reasonable hypothesis would be that training social interpretation bias in adolescents would be more effective at modifying memory bias, than training non-social situations. We could further hypothesise that modifying interpretation bias would influence mental health via shifting memory bias. Future work might seek to test this hypothesis directly with an experimental paradigm, or with extensive longitudinal data.

Network approaches also enable us to address CCBH *predictive magnitude* questions. These questions are concerned with whether cognitive biases have additive and/or interactive effects on mental health. For example, previous research has found that combined cognitive biases were more predictive of adolescent depression severity than individual biases alone (Orchard & Reynolds, 2018). Network theory, applied to mental disorders, proposes that highly connected networks are more vulnerable to psychopathology (Borsboom, 2017; also see Kalish et al. 2019). Strongly interconnected, causally related symptoms may reinforce one another to propagate a disorder. Applying network approaches to the CCBH allows us to go beyond additive and interactive effects on mental health. Specifically, our moderated network model suggests that greater connectivity amongst biases relates to lower mental health. As we have explored in this paper, network approaches allow us to examining the role of interconnectedness amongst cognitive biases (or symptoms) and relate that structure to mental health.

**Limitations**

We note several limitations of this study. We were unable to include attention bias indices in our models. The internal consistencies were so low that we would be unable to make inferences using these measures. We were therefore unable to examine the CCBH as is it usually formulated; including attention, interpretation, and memory bias indices. Although we were unaware of psychometric issues with the dot-probe at the start of this study, it is becoming clear that the task is likely unsuitable for individual differences research (for a summary, see Parsons, Kruijt, & Fox, 2019). In relation to this, our interpretation bias indices did not show optimal reliability. It is possible that this is partly due to the small number of items (5 per bias) that were presented in this task. Similarly, we do not know how reliable the memory bias task is (though the test-retest is good, which is promising). Moving forward, an important challenge for researchers in this field is to invest more time and resource into developing valid and reliable cognitive bias tasks. Otherwise, low reliability will render many CCBH questions unanswerable.

Another important issue is that our network is likely missing important mental health variables. The explained variance of mental health was only 33%. Thus, up to 67% of the variance would explained by other variables not included in the model, including; other self-reported mental health factors and life events. We opted to include only cognitive bias measures and mental health in our analyses as this was our primary interest of the present paper. We did not include any of the other psychological, social, or cognitive factors thought to comprise positive mental health (Keyes, 2002; 2005). Future work would benefit from selecting measures based on cognitive models of mental health, including symptom-level psychological outcomes. This would provide a more comprehensive test of the CCBH, and we expect would explain much more variance in the model. We also note that explained variance is limited by the reliability of the measures included, further reinforcing the need for greater psychometric scrutiny of cognitive bias measures.

**Future directions**

Psychological Network approaches offer a rapidly developing set of tools for examining the complex interplay amongst symptoms. Indeed, during the revision of this paper, moderated network models were introduced (Haslbeck, Borsboom, & Waldorp, 2019). This prompted us to reanalyse our data and use the high-low mental health network comparison as a starting point, instead of being our core analysis. Moving forward, we will be able to utilise the full three waves of data in the CogBIAS study (Booth et al. 2017). Longitudinal data offers the opportunity to examine the stability of the baseline networks presented in this paper, through adolescence. One study found that symptom networks were related to the longitudinal course of depression; more densely connected networks were associated with persistinant major depressive disorder two years later (van Borkulo et al., 2015). Using a similar approach, we will be able to test whether increased network connectivity at baseline, in early adolescence, predicts poorer mental health in later adolescence. Using all three waves of data we will also be able to use cross-lagged network models (Rhemtulla, van Bork, & Cramer, 2019) to model longitudinal changes in the cognitive bias network. We will also be able to examine whether the strength of network connectivity predicts the future network structure. For example, we might hypothesise that a denser network would become denser over time, as biases reinforce one another. Whereas a sparse network might be expected to remain sparse, as biases remain relatively independent. Longitudinal network approaches offer the opportunity to model complex interactions amongst cognitive biases over time, and therefore, examine the CCBH in greater detail.

To summarise, we applied a moderated network approach to examine the interconnections amongst cognitive biases in a large sample of adolescents. To our knowledge, this is the first empirical study to report moderated network models (Haslbeck et al., 2019). We hope to have shown the usefulness of a moderated network approach in moving beyond a static structure of a symptom network, and to examine mental health related changes in the structure. Network analyses offer a valuable tool in examining the Combined Cognitive Bias Hypothesis and an approach incorporating the complexity of interacting cognitive biases.

**References**

Rhemtulla, M., van Bork, R., & Cramer, A. O. J. (Accepted/In press). Cross-lagged network models. *Multivariate Behavioral Research*.

Notes / bits and pieces from elsewhere

**From Jonas on justifying networks and MNM.**

However, I think there are other strategies to convince the reader/reviewer that you are not doing something weird when using those models. For example, you can point out that network models can be estimated via nodewise regression, which means that from a statistical perspective, moderated network models are in the end just a set of regressions with moderation (interaction) terms. And nobody would argue with using moderation/interaction terms in regression models. You can also say that many people in the network literature test for group differences, usually with the Network Comparison Test (NCT) or the Fused Graphical Lasso (FGL). In a way those are special cases of testing moderation for a binary moderator.

**Cut from elsewhere**

These results provide some indication that biased cognition in one domain may not act in isolation, but rather may influence other processing biases to ultimately influence emotional vulnerability.

The *complexity* principle proposes that mental disorders are characterised in terms of the interactions among different components in a psychopathology network. That is to say that while the individual symptoms influence mental disorders, such symptoms are also related to one another. This relates to the *direct causal connections* principle, which states that direct causal connections among symptoms form the network structure.

Trait happiness and satisfaction with life, for example, which are both core components of positive mental health, have been associated with attentional biases towards positive images especially in the later stages of passive viewing (Raila, Scholl, & Gruber, 2015).

**Somewhat might use maybe**

Adolescence is a period of fast social and interpersonal change and development (e.g. Nelson, Leibenluft, McClure, & Pine, 2005) and it is no surprise that the interpretation of social scenarios plays a central role in adolescent mental health.

This is encouraging, and is in line with recommendations for preventative interventions to target the most at-risk samples (e.g. Horowitz & Garber, 2006) and targeting a specific risk factor (e.g. Topper, Emmelkamp, Watkins, & Ehring, 2017)

I think you need a paragraph about the actual bias variables and what their connectivity means. The parts I find really interesting are (1) greater connectivity in low MH group – does this reflect fixed cognition? Could bring in discussion on flexibility and resilience perhaps. (2) Memory bias was most strongly associated with mental health – this matches our other papers/results, (3) all biases were connected, with social interpretation bias in the centre? – mention predictability metric (4) some more about the interaction stuff, which I’m not fully understanding.

An important recent development is the recent introduction of hybrid symptom-and-resilience-factor networks to examine resilience (Kalish et al. 2019). Resilience factors may act as moderators of symptom-symptom relations, or of symptom autoconnections (i.e. a symptom predicting itself at the subsequent timepoint). This novel approach is mostly conceptual and will be subject to strong methodological demands.

Yet, this advancement holds promise for examining the CCBH.