

Published in final edited form as:

J Affect Disord. 2021 March 01; 282: 299–307. doi:10.1016/j.jad.2020.12.133.

Recall bias during adolescence: Gender differences and associations with depressive symptoms

Jessica K. Bone^{a,*}, Gemma Lewis^a, Jonathan P. Roiser^b, Sarah-Jayne Blakemore^{b,c}, Glyn Lewis^a

^aDivision of Psychiatry, University College London, London, UK

^bInstitute of Cognitive Neuroscience, University College London, London, UK

^cDepartment of Psychology, University of Cambridge, Cambridge, UK

Introduction

Throughout adulthood, women are twice as likely to experience depression as men (Salk et al., 2017). This gender difference emerges due to a sharp increase in the incidence of depression in girls relative to boys during mid-adolescence (Hankin et al., 1998; Kwong et al., 2019). We do not understand why this increase in depression occurs but, in order to prevent it, we must identify modifiable risk factors.

According to classic cognitive models of depression, depressed individuals have negative thoughts and beliefs about themselves and the world, which result from early experiences and influence information processing (how you interpret, learn from, and remember your environment; Beck, 2008, 1979; Roiser et al., 2012). Consistent with this, there is evidence that people with depression have reduced positive, or increased negative, information processing, measured using cognitive tasks (Beck, 2008; Roiser and Sahakian, 2017).

Socialisation and gender inequalities during childhood may cause girls to have more negative thoughts and beliefs about themselves (self-schema), and more negative information processing, compared to boys (Bone et al., 2020). Negative self-schema and information processing biases may be more prevalent in girls from early adolescence and contribute to the increase in the incidence of depression (Bone et al., 2020). Alternatively, negative information biases may become more prevalent in girls from mid-adolescence, alongside increases in depression.

Memory is an important aspect of information processing. Self-referential memory is usually tested by asking individuals to rate whether positive and negative personality characteristics describe themselves, followed by a surprise recall test in which participants are asked to remember as many characteristics as possible. Recall biases may be consistent with self-schema, as information about the self is preferentially remembered compared to information about others (the self-reference effect; Rogers et al., 1977; Symons and

*Corresponding author at: UCL Division of Psychiatry, 6th Floor, Maple House, 149 Tottenham Court Road, London, W1T 7BN, UK. jessica.bone@ucl.ac.uk (J.K. Bone).

Johnson, 1997). Healthy adolescents generally recall more positive than negative self-referential information, which may reduce their risk of depression (Auerbach et al., 2016; Cole et al., 2014; Connolly et al., 2016; Dainer-Best et al., 2018; Fattahi Asl et al., 2015; Hammen and Zupan, 1984; Kuiper and MacDonald, 1982; Prieto et al., 1992; Taylor and Ingram, 1999; Timbremont and Braet, 2004).

Self-referential recall may be particularly important in adolescence because the social self-concept develops during this period. Adolescents become more aware of, and concerned with, other people's opinions of them (Parker et al., 2006; Sebastian et al., 2008). Self-evaluations become more negative and self-esteem declines sharply, particularly in girls (Robins and Trzesniewski, 2005; van der Aar et al., 2018). Negative self-referential recall biases may lead to increased depressive symptoms in adolescence, and this risk factor may be more prevalent in girls (Bone et al., 2020). It is unclear whether this risk factor would be present from early adolescence or emerge during adolescence.

However, a review did not find consistent evidence of recall biases in adolescent depression (Platt et al., 2017). Depressive symptoms have been associated with poorer recall of positive information, greater recall of negative information, or a combination of both biases (Alloy et al., 2012; Asarnow et al., 2014; Fattahi Asl et al., 2015; Gençöz et al., 2001; Orchard and Reynolds, 2018; Speed et al., 2016; Woolgar and Tranah, 2010). Others have found no evidence for an association between recall biases and depressive symptoms (Dainer-Best et al., 2018; Holt et al., 2016; Reid et al., 2006). The inconsistent evidence may be due to methodological limitations. Many studies use small samples and divide participants into groups according to presence of symptoms or risk of depression, limiting statistical power. It is generally accepted that depression is a continuum (Hankin et al., 2005). Using depressive symptoms continuously in analyses should increase the sensitivity to detect any associations with recall.

Very few studies have tested whether negative recall biases are more prevalent in girls during adolescence. A longitudinal cohort study found evidence that girls had more positive recall than boys around age 13, but there were no gender differences in negative recall (McArthur et al., 2019). Changes in recall bias (from 13 to 19 years) did not differ according to gender. This study did not measure depressive symptoms so could not test whether they were associated with recall bias (McArthur et al., 2019).

In this study, we addressed these issues using a novel recall task in a large cross-sectional study (n=578). Adolescents were recruited from two age groups (young adolescents aged 11-12 years, mid-adolescents aged 13-15 years) to study recall biases before and after the gender difference in depression begins to emerge (Kwong et al., 2019). Depressive symptoms ranged from mild to severe. As previous findings with the traditional recall task are inconsistent, we developed a novel test of recall of social evaluation. Social evaluation was positive and negative personality traits, seen in a task where participants learned whether they or another person were liked or disliked. We measured recall of positive and negative words which were seen describing the self (self-referential) or another person (other-referential).

We aimed to test whether negative biases in recalling self-referential social evaluation were more prevalent in girls and were associated with depressive symptoms. To do this, we tested hypotheses relating to overall recall biases, gender differences, and associations with depressive symptoms. We hypothesised that, overall, adolescents would have a positive self-referential bias, recalling more self-referential than other-referential words (hypothesis 1), and more self-referential positive than self-referential negative words (hypothesis 2). We expected girls to demonstrate more negative self-referential recall biases than boys, recalling fewer self-referential positive and more negative words (hypothesis 3). We hypothesised that this gender difference would be present from early adolescence, so would not differ across age groups (hypothesis 4). We also expected positive self-referential recall biases to be negatively associated with depressive symptoms (hypothesis 5). Finally, we hypothesised that this association with depressive symptoms would be consistent across genders and age groups (hypothesis 6).

Methods

Participants

Participants were recruited from two age groups, Year 7 (11-12 years old) and Years 9-10 (13-15 years old), from eight diverse mixed-gender secondary schools across London. We sampled from two separate groups, maximising power to test gender differences before and after the age at which rates of depression start increasing (Kwong et al., 2019). To show a minimal difference of 0.4 standard deviations in recall between males and females ($\alpha=0.05$, power=80%), a sample of 320 adolescents was needed. To test gender differences within both age groups, we aimed to recruit 640 participants in total. There were no restrictions on whether adolescents had any mental or physical health problems or were receiving psychotropic medication or psychological therapy.

Parental consent was provided for 687 adolescents. Parental consent rates varied from 7% to 98% in each school ($M=39%$, $SD=33%$). Of these adolescents, 606 (88%) provided informed assent. Data on depressive symptoms and/or the recall task was missing for 28 (5%) participants (final $n=578$). Overall, 141 (24%) adolescents were recruited from five schools with low parental consent (under 30%) and 437 (76%) adolescents were recruited from three schools with high parental consent (over 60%). Schools with low versus high consent did not differ in age, gender, or depressive symptoms ($p>0.05$). However, participants from schools with low parental consent had higher non-verbal IQ score (mean diff=1.25, 95% CI=0.88 to 1.63, $p<0.001$) and better recall (total hits mean diff=2.89, 95% CI=2.00 to 3.79, $p<0.001$).

Ethical approval

Ethical approval was obtained from University College London (project 3453/001). Informed consent was provided by all parents/carers of participants and informed assent was provided by all participants. Participants' parents provided informed opt-in or opt-out consent, dependent on the school their child was attending. Only seven parents chose to opt-out (2% of those contacted). All procedures complied with the ethical standards of the

relevant committees on human experimentation, the Helsinki Declaration (2008 revision), and the General Data Protection Regulation.

Participants could opt-in to a prize draw for a £50 Amazon voucher after completing questionnaires at home.

Measures

Surprise recall test—Incidental memory was assessed using a surprise recall test. This differed to previous tests which ask participants whether personality characteristics describe themselves, then measure recall of words classified as self-referential. In contrast, in this novel task, we tested recall of personality descriptors previously seen in a social evaluation learning task. This method differentiated recall of self-referential and other-referential information from social interactions. This allowed us to test if recall of all social evaluation was associated with gender and depressive symptoms, or whether associations were specific self-referential information.

The social evaluation learning task was a two-alternative forced choice task based on probabilistic stimulus-reward learning tasks (Button et al., 2015). Participants learnt whether a person was liked or disliked by a computer character. Learning occurred in two conditions: about the participant themselves (self-referential) or about another person (other-referential). On each trial, a positive and negative word pair was presented (e.g. funny/grumpy). Participants were asked to choose the word which best corresponded to what the character thought about them (self-referential) or the other person (other-referential). Participants received probabilistic feedback about whether this choice was correct and used trial and error to learn whether the character liked or disliked them (or the other person) over 20 trials (Figure 1; feedback contingency 80%). For each character, one of two social rules was learnt: person is liked or disliked by the character. There were thus four blocks in this task: self-like, self-dislike, other-like, and other-dislike. Twenty word pairs were seen for the self, and 20 for the other person, with each word pair seen twice (once each in the like and dislike blocks). Performance on this task will be modelled and published separately.

Personality descriptors were emotive adjectives describing trait characteristics (e.g. cool/boring, funny/grumpy, generous/greedy). Positive and negative words were selected from databases according to their age of acquisition (Brysbaert and New, 2009; Grünh, 2016; Ku era and Francis, 1967; Leech et al., 2014; Warriner et al., 2013). The oldest mean age of acquisition of any included word was 8.78 years ($SD=1.99$). Positive and negative words were paired, matched firstly on age of acquisition. We also aimed to pair semantically linked words, minimise differences in psycholinguistic parameters (number of syllables, usage frequency, meaningfulness, familiarity, arousal), and maximise differences in likeableness, valence, and desirability ratings.

After a delay of approximately 4mins participants were asked to remember as many personality descriptors as possible in the surprise recall task. They were given 2mins to type responses. A countdown timer appeared for the final 30s. See Supplementary Figure 1 for further details. Misspelled words that resembled correct responses were categorised as correct to ensure that spelling errors did not bias accuracy. Number of self-referential

and other-referential positive and negative words accurately recalled (hits), and positive and negative incorrect responses (false alarms) were calculated.

Depressive symptoms—Participants completed the Mood and Feelings Questionnaire (short version; SMFQ), a 13-item self-report measure of depressive symptoms over the last two weeks (Angold et al., 1995). Items were rated on a scale of 0-2 (total 0-26), with higher scores indicating greater severity. Although the SMFQ is not a diagnostic measure, scores of 12 or higher may indicate the presence of depression. Missing responses were imputed for participants who responded to 10 or more questions using each individual's mean SMFQ score ($n=111$, 19%).

Confounders—Participants completed an abbreviated nine-item version of the Raven Standard Progressive Matrices Test (non-verbal IQ score; Bilker et al., 2012). Additional confounders were collected through a parental questionnaire. All parents were asked to complete this questionnaire, but response rates were low ($n=340$, 59%). Analyses were repeated controlling for additional confounders (ethnicity, English as a first language, dyslexia, autism spectrum disorders, parental education, maternal depression, paternal depression) in the Supplement.

We also intended to adjust for pubertal stage. Following classroom data collection, participants were asked to complete the Pubertal Development Scale (PDS; Petersen et al., 1988) at home. Only 117 (20%) participants completed the PDS. Analyses including pubertal stage are in the Supplement.

Procedure

Data collection was computerised and completed online using Gorilla (www.gorilla.sc). It took place with groups of 2-31 adolescents in classrooms using computers, laptops, or tablets. After providing informed assent, participants completed a battery of measures, intended for use in several studies. Participants first completed the social evaluation learning task, followed by the Raven Standard Progressive Matrices Test, and the surprise recall task. The SMFQ was then completed, followed by other questionnaires (Affective Reactivity Index, Revised Children's Anxiety and Depression Scale, Dysfunctional Attitude Scale, Health and Social Risks Questionnaire, Adolescent Social Reward Questionnaire, Children's Rejection Sensitivity Questionnaire). After classroom data collection, participants were emailed a link to complete questionnaires with more sensitive content at home (PDS, Strengths and Difficulties Questionnaire, Olweus Bully/Victim Questionnaire). We decided a priori which measures would be analysed in this study.

Statistical analyses

Analyses were performed using Stata 16 (StataCorp, 2019). As we aimed to compare the influence of gender in each age group, all descriptive statistics were presented separately for each subgroup. At this stage, we found no evidence that false alarms differed according to age group or gender, and false alarms were not associated with depressive symptoms, so they were not analysed further (Table 1).

Negative binomial mixed models—There were four types of hits: self-referential positive, self-referential negative, other-referential positive, and other-referential negative. In order to analyse such data, analysis of variance (ordinary least squares, under Gaussian assumptions) would often be used, testing whether recall differed according to various factors. However, given that hits were count variables which were positively skewed and over-dispersed, negative binomial mixed models were considered more appropriate. The four types of recall were clustered within each individual, with total number of hits as the dependent variable, and a random intercept for participant to account for clustering. Task conditions (self-/other-referential, positive/negative), gender, age group, and confounders (continuous age within each age group, school, testing group size, non-verbal IQ score, and positive and negative false alarms) were estimated as fixed effects.

We used negative binomial mixed models to calculate a hits ratio as the effect estimate, which represents the number of hits in one category relative to another (e.g. the ratio of negative to positive hits). A hits ratio larger than one meant that hits were lower in the reference category (e.g. more negative than positive hits). All models are presented before and after adjustment for confounders.

Recall biases—Our first question was whether hits differed according to word valence (positive/negative) and the condition in which words were learned (self-referential/other-referential; hypothesis 1). We included condition and valence as independent variables with hits as the dependent variable. Next, we added an interaction between condition and valence, to test whether the association between valence and recall differed for self-referential versus other-referential words (hypothesis 2).

Gender differences—Next, we examined gender differences in recall (hypothesis 3). We tested a three-way interaction between gender, condition and valence with hits as the dependent variable, and report the two-way interactions between these variables. To assess whether gender differences were consistent across age groups, we tested a four-way interaction between age group, gender, condition and valence with hits as the dependent variable (hypothesis 4). As our aim was to compare the influence of gender in each age group, we only report the lower level (two-way and three-way) interactions which include gender. Where there was evidence of an interaction, we examined associations with hits separately for each subgroup. Additionally, to check that gender differences in hits were not explained by depressive symptoms, we added depressive symptoms to the negative binomial mixed models (Supplement).

Associations with depressive symptoms—Finally, we examined whether recall was associated with depressive symptoms (hypothesis 5). Linear regression tested whether self-referential positive, self-referential negative, other-referential positive, and other-referential negative hits were associated with depressive symptoms (SMFQ score; continuous dependent variable). For this analysis, all task parameters were included in a single model to adjust for overall performance. This model was adjusted for age group and gender in addition to other confounders.

For each type of hit associated with depressive symptoms, we tested whether the association differed according to age group and gender (hypothesis 6). We added a three-way interaction between hits, age group, and gender to the linear regression model with depressive symptoms as the dependent variable. We also included two-way interactions between hits and age group, and hits and gender.

Results

Our sample consisted of 578 adolescents (49% female). 315 (54%) were young adolescents from Year 7, and 263 (46%) were mid-adolescents from Years 9-10. Young adolescents' age ranged from 11 to 12 years (except for one participant aged 13; $M=11.56$, $SD=0.50$) and mid-adolescents' age ranged from 13 to 15 years ($M=14.18$, $SD=0.51$). Table 1 shows characteristics and task performance according to age group and gender.

In young adolescents, SMFQ score ranged from 0 to 23 ($M=7.14$, $SD=5.49$). The SMFQ threshold for depression was met by 60 (18%; 58% of whom were female) young adolescents. In mid-adolescents, SMFQ score ranged from 0 to 26 ($M=8.27$, $SD=5.86$). The SMFQ threshold for depression was met by 61 (23%; 67% of whom were female) mid-adolescents.

There was no evidence that depressive symptoms were higher in mid-than young adolescents (adjusted $p=0.39$), but depressive symptoms were higher in females than males (adjusted $\text{coef}=2.28$, 95% $CI=1.34$ to 3.23 , $p<0.001$). There was very weak evidence of an interaction between age group and gender on depressive symptoms (adjusted $p=0.08$). The gender difference in depressive symptoms increased with age (young adolescents $\text{coef}=1.65$, 95% $CI=0.40$ to 2.90 ; mid-adolescents $\text{coef}=3.10$, 95% $CI=1.69$ to 4.51 adjusted).

Recall biases

Hypothesis 1: Participants made 7% more self-referential than other-referential hits (adjusted hits ratio= 1.07 , 95% $CI=1.02$ to 1.13 , $p=0.01$) and 14% more negative than positive hits (adjusted hits ratio= 1.14 , 95% $CI=1.08$ to 1.20 , $p<0.001$).

Hypothesis 2: There was no evidence for an interaction between valence and condition (adjusted $p=0.25$). Participants made more negative than positive self-referential and other-referential hits (Table 1).

Gender differences

Overall, there was evidence that females made 18% more hits than males (95% $CI=1.09$ to 1.28 , $p<0.001$ adjusted) and mid-adolescents made 43% more hits than young adolescents (95% $CI=1.13$ to 1.81 , $p=0.003$ adjusted).

Hypothesis 3: There was evidence for a two-way interaction between gender and condition (adjusted $p=0.04$). Males made more self-referential than other referential hits, whereas females did not show this self-reference effect (Table 2). There was no evidence for a two-way interaction between gender and valence (adjusted $p=0.99$), as both males and females made more negative than positive hits (Table 2). There was also no evidence for a

three-way interaction between gender, condition, and valence on hits (adjusted $p=0.87$; Table 2).

Hypothesis 4: Next, we tested whether the gender differences in recall differed across age groups. There was no evidence for a two-way interaction between age group and gender on total hits (adjusted $p=0.10$). In both age groups, females made more hits than males (Table 2). However, there was weak evidence for a three-way interaction between age group, gender and condition (adjusted $p=0.05$). Young adolescent females made slightly fewer self-referential than other-referential hits but, in all other groups, more self-referential than other-referential hits were made (Table 2; Figure 2). There was no evidence that the number of positive versus negative hits differed according to age group and gender (adjusted $p=0.43$). All groups made more negative than positive hits (Table 2; Figure 2). Finally, there was no evidence of a four-way interaction between age group, gender, condition and valence (adjusted $p=0.42$; Figure 2).

Adjusting for depressive symptoms did not substantially alter these findings (Supplement).

Associations with depressive symptoms

Hypothesis 5: There was evidence for an association between positive and negative self-referential hits and depressive symptoms (Table 3). For each additional self-referential negative hit, SMFQ score increased by 0.45 points (95% CI=0.15 to 0.75, $p=0.003$ adjusted). In contrast, self-referential positive hits were negatively associated with depressive symptoms. For each additional self-referential positive hit, SMFQ score decreased by 0.38 points (95% CI=-0.69 to -0.08, $p=0.01$ adjusted).

Before adjusting for confounders, there was evidence that other-referential negative hits were positively associated with depressive symptoms (coef=0.34, 95% CI=0.05 to 0.62, $p=0.02$). However, this no longer achieved significance after adjustment for confounders (coef=0.24, 95% CI=-0.05 to 0.54, $p=0.11$ adjusted). There was no evidence that other-referential positive hits were associated with depressive symptoms (Table 3).

Hypothesis 6: Associations between self-referential hits and depressive symptoms did not differ across age groups (adjusted interactions: positive $p=0.57$; negative $p=0.41$). The association between self-referential positive hits and depressive symptoms also did not differ according to gender (adjusted interaction $p=0.47$). However, the association between self-referential negative hits and depressive symptoms was larger in females (adjusted coef=0.85, 95% CI=0.36 to 1.34) than in males (adjusted coef=0.27, 95% CI=-0.13 to 0.67). There was weak evidence for this interaction (adjusted $p=0.04$). This gender difference in the association between self-referential negative hits and depressive symptoms was present across age groups. There was no evidence for three-way interactions between age group, gender and self-referential hits on depressive symptoms (adjusted: positive $p=0.52$; negative $p=0.30$).

Discussion

Following a social evaluation learning task, adolescents were asked to recall this self-referential and other-referential social evaluation. Consistent with our first hypothesis, most adolescents better recalled self-referential than other-referential words, demonstrating a self-referential bias. However, young adolescent girls (11-12 years) recalled fewer self-referential than other-referential words, which was unexpected. We hypothesised that adolescents' self-referential bias would be positive, with better recall of self-referential positive than self-referential negative words (hypothesis 2). However, adolescents recalled more negative than positive words in both self-referential and other-referential conditions. Although we expected girls to demonstrate more negative self-referential recall biases than boys (hypothesis 3), we found no other evidence for gender differences in recall in either age group (hypothesis 4).

As predicted in hypothesis 5, more severe depressive symptoms were associated with a decrease in self-referential positive recall and an increase in self-referential negative recall. These associations were similar across early and mid-adolescence, as outlined in hypothesis 6. However, contrary to hypothesis 6, the association between self-referential negative recall and depressive symptoms was more pronounced in girls than boys.

We found evidence of enhanced memory for self-referential information, as previously shown in children (Cunningham et al., 2014) and adults (Symons and Johnson, 1997). It is unclear why this self-reference effect was not present in young adolescent girls. It is possible that this group found social evaluation about others more salient or paid more attention to other-referential evaluation, and thus better remembered other-referential words, compared to other adolescents. However, we do not have any evidence to support this explanation.

In contrast to previous studies with healthy adults (Denny and Hunt, 1992; Sanz, 1996; Sedikides and Green, 2000) and adolescents (Auerbach et al., 2016; Cole et al., 2014; Connolly et al., 2016; Dainer-Best et al., 2018; Fattahi Asl et al., 2015; Hammen and Zupan, 1984; Kuiper and MacDonald, 1982; Prieto et al., 1992; Taylor and Ingram, 1999; Timbremont and Braet, 2004), we did not find evidence for positively biased self-referential recall. Adolescents recalled more negative than positive words in all conditions. This may be because self-evaluations become more negative and self-esteem declines during adolescence (Robins and Trzesniewski, 2005; van der Aar et al., 2018). However, this would account for biases only in *self*-referential recall. The generalisation of this negative bias to other-referential recall could be due to our encoding task. Words were viewed as social evaluation, which may make negative words more salient and boost memory, regardless of whether words refer to the self or others. Consistent with this explanation, another study using a social evaluative encoding task (participants imagined overhearing others describing them) also found that adolescents remembered more negative than positive words (Holt et al., 2016).

Contrary to our hypotheses, we did not find a gender difference in negative recall biases. We have previously proposed that gender inequality may cause girls to have more negative self-schema, which could lead to more negative recall biases (Bone et al., 2020). It is

possible that girls have more negative self-schema in adolescence, but this was not captured by performance on our surprise recall task. However, the only previous study of gender differences in recall biases during adolescence found evidence that girls had more positive recall than boys, and there were no gender differences in negative recall (McArthur et al., 2019). This is opposite to the gender difference that we proposed. Therefore, despite the evidence for an association between negative recall biases and depressive symptoms, and more severe depressive symptoms in girls than boys, girls may not have more negative recall biases during adolescence. We did find some evidence that the association between self-referential negative recall and depressive symptoms was stronger in girls than boys across age groups. This was unexpected as we anticipated that recall would be similarly associated with depressive symptoms across genders. If self-referential negative recall is a risk factor for depressive symptoms, it may be more important for girls.

Both increased negative and reduced positive self-referential recall were associated with depressive symptoms, as found in some previous studies (Fattahi Asl et al., 2015; Speed et al., 2016). This finding differs to a recent review, which did not find consistent evidence for memory biases in adolescent depression (Platt et al., 2017). This could be because previous studies have generally assessed the proportion of words previously endorsed as self-referential that are recalled. Testing recall of social evaluation may provide a more nuanced measure of memory biases.

In this study, effect estimates and confidence intervals for the associations between self-referential positive hits and depressive symptoms were clearly different from the corresponding association with other-referential positive hits, potentially suggesting a specific role of poorer self-referential positive recall in vulnerability to depressive symptoms. It is less clear whether there is a specific role of self-referential negative recall. In unadjusted analyses, self-referential and other-referential negative hits were associated with depressive symptoms. After adjusting for confounders, evidence for the association between other-referential negative hits and depressive symptoms was attenuated, but the coefficient and confidence interval were not clearly different from those for the corresponding association with self-referential negative hits. We cannot rule out that the association between self-referential negative hits and depressive symptoms reflects a general negative bias. However, self-referential negative recall was most strongly associated with depressive symptom severity, as previously found (Dainer-Best et al., 2018).

Strengths and limitations

We aimed to test whether gender differences in recall bias were associated with depressive symptoms in adolescence. Our sample was population-based and included the full range of depressive symptoms (from none to severe), which we analysed continuously. This should have increased our statistical power to detect any associations between recall bias and depressive symptoms (Button et al., 2013), although this study may have been underpowered for testing three- and four-way interactions. The sample was recruited from eight diverse schools, making it more representative than many previous studies. We used a novel recall task, allowing us to differentiate self-referential and other-referential recall bias.

However, this recall task had some limitations. Its reliability and validity are unknown, although tasks assessing memory and emotional biases are generally reliable (Bland et al., 2016). The nature of the encoding task may have influenced recall. Traditional tasks measure recall of words describing how participants see themselves, rather than how another individual sees them. Whilst information consistent with the self-concept was probably preferentially recalled, words incongruent with the self-concept may have been more memorable. Adolescents with more depressive symptoms could have been differentially affected by the idea of someone liking or disliking them, altering reactions to the words, and influencing recall.

The poor parental consent rates in several schools was a limitation. Selection bias may have occurred, as participants had higher non-verbal IQ and better recall in schools with low parental consent. However, 76% of the sample were from schools with high consent. We do not think that the factors influencing selection bias would alter associations between recall bias and depressive symptoms. Opt-out consent was used to recruit nearly half of our sample, which should also have reduced selection bias.

Although we adjusted for several potential confounders, residual confounding is also possible. For subsamples with information available on additional confounders (59% of participants) and pubertal stage (20% of participants), adjusting for these potential confounders did not alter the evidence for any associations (Supplement). However, in these subsamples, there was no evidence for associations between self-referential positive hits and depressive symptoms. Nevertheless, the effect estimates were similar to the coefficients, and within the confidence intervals, from the primary analyses with the whole sample. The lack of evidence could be due to the reduced sample size or selection bias in participants with data on these potential confounders.

As this was a cross-sectional study, we cannot provide evidence of a causal effect of recall bias on depressive symptoms, as proposed by cognitive models of depression (Roiser and Sahakian, 2017). Our findings are consistent with such models. However, it is equally possible that changes in depressive symptoms cause changes in recall biases (reverse causality), or that the association is bidirectional. Longitudinal data is required to test the hypothesis that negatively biased recall leads to increased depressive symptoms.

Consistent with contemporary cognitive models of depression (Beck, 2008; Roiser and Sahakian, 2017), adolescents who had more self-referential negative and less self-referential positive recall had more severe depressive symptoms. There was no evidence for gender or age differences in these recall biases, although there was some evidence that the association with self-referential negative recall was stronger in girls. The association between recall biases and depressive symptoms was similar across early and mid-adolescence, despite the increase in depressive symptoms in older adolescents. Negatively biased self-referential recall may lead to more negative memories of social interactions and more negative self-concepts, encouraging social withdrawal and increasing depressive symptoms. Negative self-referential bias may be a risk factor for the emergence of depressive symptoms during adolescence and is a good candidate for future longitudinal studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Alloy LB, Black SK, Young ME, Goldstein KE, Shapero BG, Stange JP, Boccia AS, Matt LM, Boland EM, Moore LC, Abramson LY. Cognitive vulnerabilities and depression versus other psychopathology symptoms and diagnoses in early adolescence. *J Clin Child Adolesc Psychol*. 2012; 41: 539–560. DOI: 10.1080/15374416.2012.703123 [PubMed: 22853629]
- Angold A, Costello EJ, Messer SC, Pickles A. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *Int J Methods Psychiatr Res*. 1995; 5: 237–249.
- Asarnow LD, Thompson RJ, Joormann J, Gotlib IH. Children at risk for depression: Memory biases, self-schemas, and genotypic variation. *J Affect Disord*. 2014; 159: 66–72. [PubMed: 24679392]
- Auerbach RP, Bondy E, Stanton CH, Webb CA, Shankman SA, Pizzagalli DA. Self-referential processing in adolescents: Stability of behavioral and ERP markers. *Psychophysiology*. 2016; 53: 1398–1406. DOI: 10.1111/psyp.12686 [PubMed: 27302282]
- Beck AT. The evolution of the cognitive model of depression and its neurobiological correlates. *Am J Psychiatry*. 2008; 165: 969–977. DOI: 10.1176/appi.ajp.2008.08050721 [PubMed: 18628348]
- Beck, AT. *Cognitive therapy of depression*. Guilford press; New York, NY: 1979.
- Bilker WB, Hansen JA, Brensinger CM, Richard J, Gur RE, Gur RC. Development of Abbreviated Nine-Item Forms of the Raven's Standard Progressive Matrices Test. *Assessment*. 2012; 19: 354–369. DOI: 10.1177/1073191112446655 [PubMed: 22605785]
- Bland AR, Roiser JP, Mehta MA, Schei T, Boland H, Campbell-Meiklejohn DK, Emsley RA, Munafo MR, Penton-Voak IS, Seara-Cardoso A, Viding E, et al. EMOTICOM: A Neuropsychological Test Battery to Evaluate Emotion, Motivation, Impulsivity, and Social Cognition. *Front Behav Neurosci*. 2016; 10: 1–17. [PubMed: 26834590]
- Bone JK, Lewis Gemma, Lewis Glyn. The role of gender inequalities in adolescent depression. *The Lancet Psychiatry*. 2020; 7: 471–472. [PubMed: 32445675]
- Brybaert M, New B. Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav Res Methods*. 2009; 41: 977–990. DOI: 10.3758/BRM.41.4.977 [PubMed: 19897807]
- Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafo MR. Confidence and precision increase with high statistical power. *Nat Rev Neurosci*. 2013; 14: 585. doi: 10.1038/nrn3475
- Button KS, Kounali DZ, Stapinski L, Rapee RM, Lewis G, Munafo MR. Fear of negative evaluation biases social evaluation inference: Evidence from a probabilistic learning task. *PLoS One*. 2015; 10: 1–15. DOI: 10.1371/journal.pone.0119456
- Cole DA, Dukewich TL, Roeder K, Sinclair KR, McMillan J, Will E, Bilsky SA, Martin NC, Felton JW. Linking peer victimization to the development of depressive self-schemas in children and adolescents. *J Abnorm Child Psychol*. 2014; 42: 149–160. DOI: 10.1007/s10802-013-9769-1 [PubMed: 23824686]
- Connolly SL, Abramson LY, Alloy LB. Information processing biases concurrently and prospectively predict depressive symptoms in adolescents: Evidence from a self-referent encoding task. *Cogn Emot*. 2016; 30: 550–560. DOI: 10.1080/02699931.2015.1010488 [PubMed: 25707445]
- Cunningham SJ, Brebner JL, Quinn F, Turk DJ. The Self-Reference Effect on Memory in Early Childhood. *Child Dev*. 2014; 85: 808–823. DOI: 10.1111/cdev.12144 [PubMed: 23888928]
- Dainer-Best J, Lee HY, Shumake JD, Yeager DS, Beevers CG. Determining optimal parameters of the self-referent encoding task: A large-scale examination of self-referent cognition and depression. *Psychol Assess*. 2018; 30: 1527–1540. DOI: 10.1037/pas0000602 [PubMed: 29878818]

- Denny EB, Hunt RR. Affective Valence and Memory in Depression: Dissociation of Recall and Fragment Completion. *J Abnorm Psychol.* 1992; 101: 575–580. DOI: 10.1037/0021-843X.101.3.575 [PubMed: 1500616]
- Fattahi Asl A, Ghanizadeh A, Mollazade J, Afkaseir A. Differences of biased recall memory for emotional information among children and adolescents of mothers with MDD, children and adolescents with MDD, and normal controls. *Psychiatry Res.* 2015; 228: 223–227. DOI: 10.1016/j.psychres.2015.04.001 [PubMed: 25998002]
- Gençöz T, Voelz ZR, Gençöz F, Pettit JW, Joiner TE. Specificity of information processing styles to depressive symptoms in youth psychiatric inpatients. *J Abnorm Child Psychol.* 2001; 29: 255–262. DOI: 10.1023/A:1010385832566 [PubMed: 11411787]
- Grühd D. An English word database of emotional terms (EMOTE). *Psychol Rep.* 2016; 119: 290–308. DOI: 10.1177/0033294116658474 [PubMed: 27401069]
- Hammen CL, Zupan BA. Self-schemas, depression, and the processing of personal information in children. *J Exp Child Psychol.* 1984; 37: 598–608. DOI: 10.1016/0022-0965(84)90079-1 [PubMed: 6747551]
- Hankin BL, Abramson LY, Moffitt TE, Silva PA, McGee R, Angell KE. Development of depression from preadolescence to young adulthood: emerging gender differences in a 10-year longitudinal study. *J Abnorm Psychol.* 1998; 107: 128–140. [PubMed: 9505045]
- Hankin BL, Fraley RC, Lahey BB, Waldman ID. Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. *J Abnorm Psychol.* 2005; 114: 96–110. DOI: 10.1037/0021-843X.114.1.96 [PubMed: 15709816]
- Holt RJ, Graham JME, Whitaker KJ, Hagan CC, Ooi C, Wilkinson PO, Van Nieuwenhuizen AO, Lennox BR, Sahakian BJ, Goodyer IM, Bullmore ET, et al. Functional MRI of emotional memory in adolescent depression. *Dev Cogn Neurosci.* 2016; 19: 31–41. [PubMed: 26802367]
- Ku era, H, Francis, WN. Computational analysis of present-day American English. Dartmouth Publishing Group; 1967.
- Kuiper NA, MacDonald MR. Self and other perception in mild depressives. *Soc Cogn.* 1982; 1: 223–239.
- Kwong ASF, Manley D, Timpson NJ, Pearson RM, Heron J, Sallis H, Stergiakouli E, Davis OSP, Leckie G. Identifying Critical Points of Trajectories of Depressive Symptoms from Childhood to Young Adulthood. *J Youth Adolesc.* 2019; 48: 815–827. [PubMed: 30671716]
- Leech, G, Rayson, P, Wilson, A. Word frequencies in written and spoken English: Based on the British National Corpus. Routledge; London: 2014.
- McArthur BA, Burke TA, Connolly SL, Olino TM, Lumley MN, Abramson LY, Alloy LB. A Longitudinal Investigation of Cognitive Self-schemas across Adolescent Development. *J Youth Adolesc.* 2019; 48: 635–647. [PubMed: 30612293]
- Orchard F, Reynolds S. The combined influence of cognitions in adolescent depression: Biases of interpretation, self-evaluation, and memory. *Br J Clin Psychol.* 2018; 57: 420–435. DOI: 10.1111/bjc.12184 [PubMed: 29799126]
- Parker, JG, Rubin, KH, Erath, SA, Wojslawowicz, JC, Buskirk, AA. Developmental Psychopathology: Theory and Methods. Cicchetti, D, Cohen, DJ, editors. Vol. 1. Wiley; New York: 2006. 96–161.
- Petersen AC, Crockett L, Richards M, Boxer A. A self-report measure of pubertal status: Reliability, validity, and initial norms. *J Youth Adolesc.* 1988; 17: 117–133. DOI: 10.1007/BF01537962 [PubMed: 24277579]
- Platt B, Waters AM, Schulte-Koerne G, Engelmann L, Salemink E. A review of cognitive biases in youth depression: attention, interpretation and memory. *Cogn Emot.* 2017; 31: 462–483. DOI: 10.1080/02699931.2015.1127215 [PubMed: 26785312]
- Prieto SL, Cole DA, Tageson CW. Depressive self-schemas in clinic and nonclinic children. *Cognit Ther Res.* 1992; 16: 521–534. DOI: 10.1007/BF01175139
- Reid SC, Salmon K, Lovibond PF. Cognitive biases in childhood anxiety, depression, and aggression: Are they pervasive or specific? *Cognit Ther Res.* 2006; 30: 531–549. DOI: 10.1007/s10608-006-9077-y

- Robins RW, Trzesniewski KH. Self-esteem development across the lifespan. *Curr Dir Psychol Sci.* 2005; 14: 158–162. DOI: 10.1111/j.0963-7214.2005.00353.x
- Rogers TB, Kuiper NA, Kirker WS. Self-reference and the encoding of personal information in depression. *J Pers Soc Psychol.* 1977; 35: 677–688. [PubMed: 909043]
- Roiser JP, Elliott R, Sahakian BJ. Cognitive mechanisms of treatment in depression. *Neuropsychopharmacology.* 2012; 37: 117–36. DOI: 10.1038/npp.2011.183 [PubMed: 21976044]
- Roiser, JP, Sahakian, BJ. *The Oxford Handbook of Mood Disorders.* DeRubeis, RJ, Strunk, DR, editors. Oxford University Press; Oxford: 2017.
- Salk RH, Hyde JS, Abramson LY. Gender Differences in Depression in Representative National Samples: Meta-Analyses of Diagnoses and Symptoms. *Psychol Bull.* 2017; 143: 783–822. [PubMed: 28447828]
- Sanz J. Memory Biases in Social Anxiety and Depression. *Cogn Emot.* 1996; 10: 87–106. DOI: 10.1080/026999396380402
- Sebastian CL, Burnett S, Blakemore S-J. Development of the self-concept during adolescence. *Trends Cogn Sci.* 2008; 12: 441–446. DOI: 10.1016/j.tics.2008.07.008 [PubMed: 18805040]
- Sedikides C, Green JD. On the self-protective nature of inconsistency-negativity management: Using the person memory paradigm to examine self-referent memory. *J Pers Soc Psychol.* 2000; 79: 906–922. DOI: 10.1037/0022-3514.79.6.906 [PubMed: 11138760]
- Speed BC, Nelson BD, Auerbach RP, Klein DN, Hajcak G. Depression Risk and Electrocortical Reactivity during Self-Referential Emotional Processing in 8 to 14 Year-Old Girls. *J Abnorm Psychol.* 2016; 125: 607–619. DOI: 10.1037/abn0000173. Depression [PubMed: 27175985]
- StataCorp. *Stata Statistical Software: Release 16.* 2019.
- Symons CS, Johnson BT. The Self-Reference Effect in Memory: A Meta-Analysis. *Psychol Bull.* 1997; 121: 371–394. [PubMed: 9136641]
- Taylor L, Ingram RE. Cognitive Reactivity and Depressotypic Information Processing in Children of Depressed Mothers. *J Abnorm Psychol.* 1999; 108: 202–210. [PubMed: 10369030]
- Timbremont B, Braet C. Cognitive vulnerability in remitted depressed children and adolescents. *Behav Res Ther.* 2004; 42: 423–437. DOI: 10.1016/S0005-7967(03)00151-7 [PubMed: 14998736]
- van der Aar LPE, Peters S, Crone EA. The development of self-views across adolescence: Investigating self-descriptions with and without social comparison using a novel experimental paradigm. *Cogn Dev.* 2018; 48: 256–270.
- Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav Res Methods.* 2013; 45: 1191–1207. DOI: 10.3758/s13428-012-0314-x [PubMed: 23404613]
- Woolgar M, Tranah T. Cognitive vulnerability to depression in young people in secure accommodation: The influence of ethnicity and current suicidal ideation. *J Adolesc.* 2010; 33: 653–661. DOI: 10.1016/j.adolescence.2009.11.005 [PubMed: 20116091]

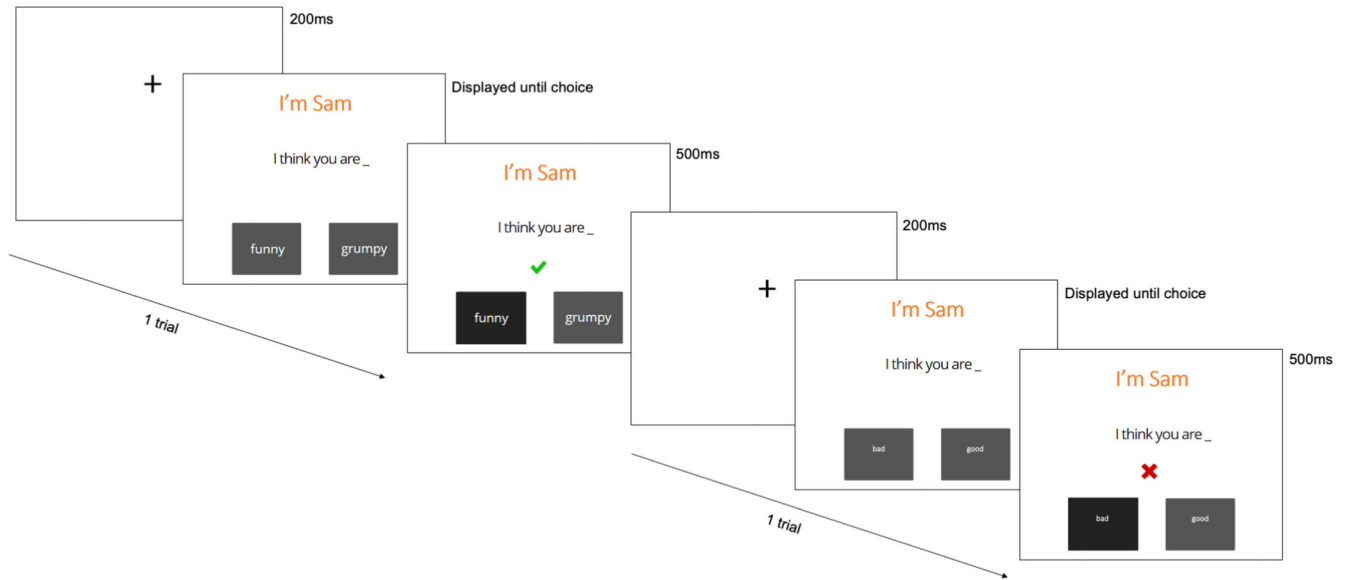


Figure 1.

Social Evaluation Learning task. An example of two trials from a self-referential block, in which the computer character is called Sam and the participant is learning what Sam thinks of them. After viewing a fixation cross, the participant was presented with a positive and negative word pair and instructed to choose the word which best corresponded with what Sam thought about them. They then received feedback about whether their choice was correct (green tick) or incorrect (red cross). Participants used trial and error to learn whether the character liked or disliked them over 20 trials. In the first trial shown here, the participant selected the positive word, which was correct. In the second trial, the participant chose the negative word, which was incorrect. Both of these trials show true (as opposed to misleading) feedback. To prevent ceiling effects, feedback contingency was set at 80%, so that 'correct' responses received an 8:2 ratio of positive to negative feedback and 'incorrect' responses received an 8:2 ratio of negative to positive feedback.

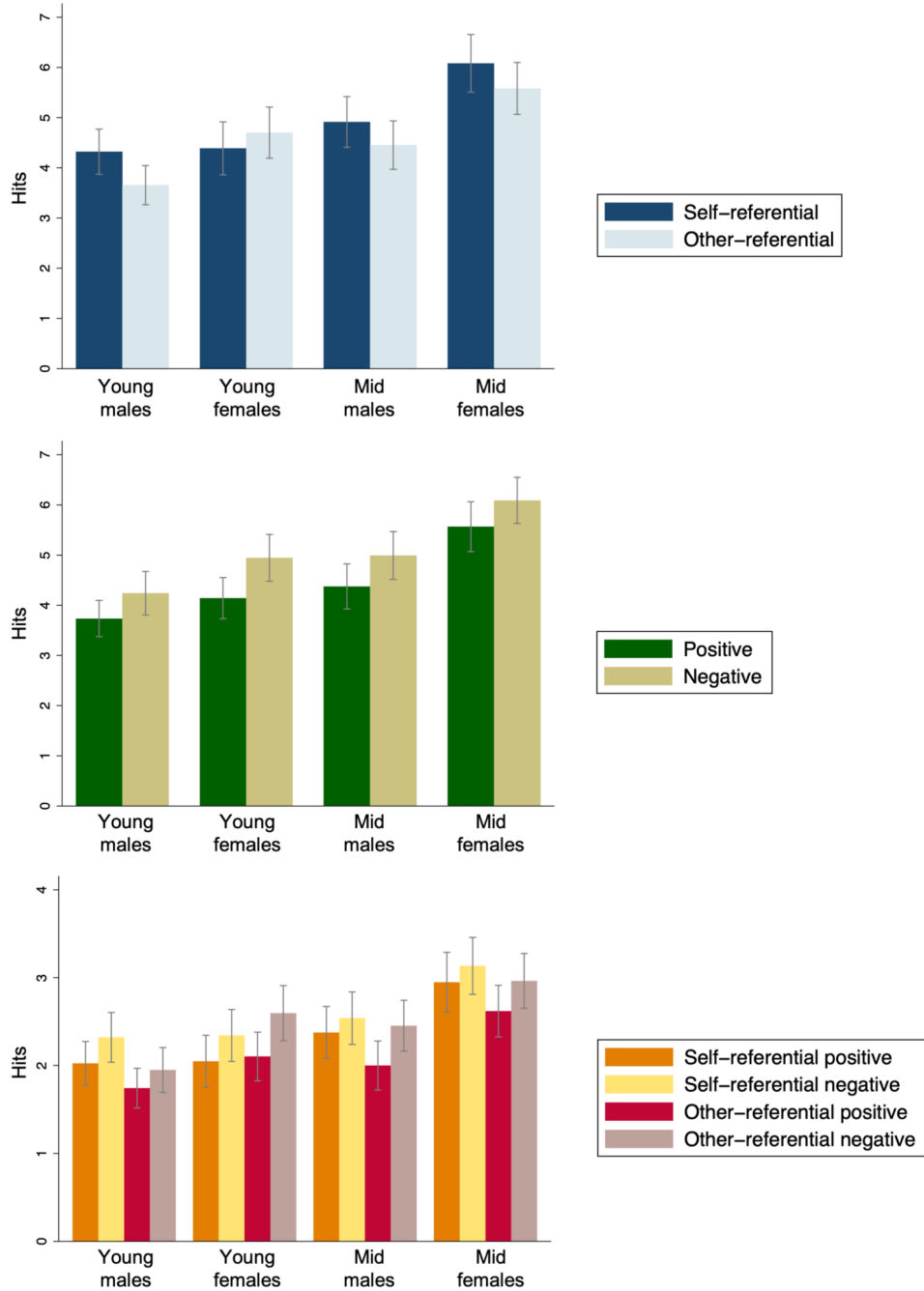


Figure 2. A) Mean hits according showing three-way interaction between age group, gender, and word condition (self-referential or other-referential). B) Mean hits showing three-way interaction between age group, gender, and word valence (positive or negative). C) Mean hits showing four-way interaction between age group, gender, condition and valence. All plotted using raw data.

Table 1
Characteristics of the full study sample according to age group and gender.

	Young adolescents		Mid-adolescents	
	Male n=159	Female n=146	Male n=128	Female n=134
Age (years)	11.56 (0.50)	11.56 (0.51)	14.20 (0.47)	14.17 (0.56)
Non-verbal IQ score	3.94 (2.02)	4.42 (1.87)	4.80 (2.03)	4.96 (2.05)
Depressive symptoms	6.47 (5.06)	7.94 (5.90)	6.69 (4.93)	9.73 (6.30)
Recall task performance				
Self-referential positive hits	2.03 (1.60)	2.05 (1.83)	2.38 (1.71)	2.95 (2.01)
Self-referential negative hits	2.32 (1.82)	2.34 (1.82)	2.54 (1.72)	3.13 (1.91)
Other-referential positive hits	1.74 (1.45)	2.10 (1.71)	2.00 (1.61)	2.62 (1.73)
Other-referential negative hits	1.95 (1.64)	2.60 (1.94)	2.45 (1.67)	2.96 (1.85)
Positive false alarms	0.55 (0.89)	0.64 (0.93)	0.69 (0.99)	0.70 (0.94)
Negative false alarms	0.67 (1.05)	0.64 (1.03)	0.59 (0.80)	0.71 (1.01)

Note. Young adolescents were recruited from Year 7 (11-12 years old) and mid-adolescents were recruited from Years 9-10 (13-15 years old). Age in years was missing for n=2 young adolescents. Gender was missing for n=10 young adolescents and n=1 mid-adolescent.

Unadjusted and adjusted negative binomial mixed models testing the effect of gender, age group, condition (whether words were learnt in relation to the self or another person) and valence (whether words were positive or negative) on the total number of hits.

Table 2

	Unadjusted models (n=567)			Fully adjusted models (n=566)		
	Interaction p value	Subgroup analysis		Interaction p value	Subgroup analysis	
		Hits ratio	95% CI		Hits ratio	95% CI
Gender x condition	0.03			0.04		
Males: condition		1.14	1.05 to 1.23		1.14	1.05 to 1.23
Females: condition		1.02	0.94 to 1.09		1.02	0.94 to 1.09
Gender x valence	1.00			0.99		
Males: valence		1.14	1.05 to 1.23		1.14	1.05 to 1.23
Females: valence		1.14	1.06 to 1.22		1.14	1.06 to 1.22
Gender x condition x valence	0.88			0.87		
Males self-referential: valence		1.11	1.00 to 1.24		1.11	1.00 to 1.24
Males other-referential: valence		1.17	1.04 to 1.31		1.17	1.04 to 1.31
Females self-referential: valence		1.10	0.99 to 1.22		1.10	0.99 to 1.22
Females other-referential: valence		1.18	1.06 to 1.31		1.18	1.06 to 1.31
Age group x gender	0.22			0.10		
Young adolescents: gender		1.13	0.99 to 1.29		1.07	0.95 to 1.20
Mid-adolescents: gender		1.26	1.12 to 1.41		1.25	1.12 to 1.40
Age group x gender x condition	0.05			0.05		
Young adolescent males: condition		1.18	1.05 to 1.31		1.18	1.05 to 1.31
Young adolescent females: condition		0.93	0.83 to 1.04		0.93	0.83 to 1.04
Mid-adolescent males: condition		1.10	0.99 to 1.24		1.10	0.99 to 1.24
Mid-adolescent females: condition		1.09	0.99 to 1.20		1.09	0.99 to 1.20
Age group x gender x valence	0.42			0.43		
Young adolescent males: valence		1.13	1.02 to 1.27		1.14	1.02 to 1.27
Young adolescent females: valence		1.19	1.06 to 1.33		1.19	1.06 to 1.33
Mid-adolescent males: valence		1.14	1.02 to 1.28		1.14	1.02 to 1.28

	Unadjusted models (n=567)			Fully adjusted models (n=566)		
	Interaction p value	Subgroup analysis		Interaction p value	Subgroup analysis	
		Hits ratio	95% CI		Hits ratio	95% CI
Mid-adolescent females: valence		1.10	0.99 to 1.21		1.10	0.99 to 1.21
Age group x gender x condition x valence	0.43			0.42		

Note. All models adjusted for condition and valence. Fully adjusted models also adjusted for continuous age within each age group, school, testing group size, non-verbal IQ score, and positive and negative false alarms. For gender, male was the reference group. For condition, other-referential was the reference group. For valence, positive was the reference group.

Table 3

Change in depressive symptoms (SMFQ score) for each additional self-referential positive, self-referential negative, other-referential positive, and other-referential negative hit.

	Model 1: unadjusted (n=578)			Model 2: fully adjusted (n=566)		
	Coef	95% CI	p value	Coef	95% CI	p value
Self-referential hits						
Positive	-0.34	-0.63 to -0.04	0.02	-0.38	-0.69 to -0.08	0.01
Negative	0.47	0.17 to 0.76	0.002	0.45	0.15 to 0.75	0.003
Other-referential hits						
Positive	0.10	-0.23 to 0.41	0.57	0.04	-0.28 to 0.37	0.79
Negative	0.34	0.05 to 0.62	0.02	0.24	-0.05 to 0.54	0.11

Note. Both models included all four types of hits as independent variables. Model 2 was adjusted for age group, gender, continuous age within each age group, school, testing group size, non-verbal IQ score, and positive and negative false alarms.