Age-related performance on an emotional Stroop task in adolescence.

Dunning, D. L.[1]*, Griffiths, K.[1]*, Parker, J.[1], Bennett, M.[1], Ahmed, S.[2], Griffin, C.[2], Foulkes, L.[3], Leung, J.T.[2], Sakhardande, A.[2], MYRIAD team[4], Kuyken, W.[4], Williams, J.M.G.[4], Blakemore, S-J.[2] & Dalgleish, T.[1,5]

*Joint first authors

Author affiliations:

[1] Medical Research Council Cognition and Brain Sciences Unit, Cambridge University, Cambridge, CB2 7EF, UK

[2] Institute of Cognitive Neuroscience, University College London, London, WC1N 3AR, UK

[3] Department of Education, University of York, York, YO10 5DD, UK

[4] Department of Psychiatry, University of Oxford, Oxford, OX3 7JX, UK

[5]Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, UK

Corresponding author:

darren.dunning@mrc-cbu.cam.ac.uk

Introduction

The Stroop task is a well-established cognitive paradigm used to measure selective attention by way of inhibitory processes. The classic version of the task involves the presentation of colour words in different coloured inks with participants asked to name the colour of the ink of each word as quickly as possible, while ignoring the meaning of the word (Stroop, 1935). Longer latencies have been reliably shown when stimuli are incongruent (e.g. when the word 'red' is printed in blue ink) rather than when they are congruent (e.g., when the word 'red' is printed in red ink). This phenomenon, known as the Stroop effect, is said to occur due to an inference in inhibiting over-learned or distracting information (MacLeod, 1991).

There are several variations on the Stroop task, the most widely used is the Emotional Stroop task (EST). The EST usually involves naming the ink colour of an emotionally valanced word (Gotlib & McCann, 1984), with the emotional Stroop effect occurring if the colour naming response is slower for an emotionally relevant word relative to a neutral word; this slower response can be seen as a reflection of a greater attentional bias toward the emotional stimuli.

Although the emotional Stroop effect has been demonstrated in a number of studies that have compared typical participants to those with high anxiety, phobia, and posttraumatic stress disorder (Bremner et al., 2004; Williams, Mathews, & MacLeod, 1996), reviews generally conclude that deficits in EST performance in depression are equivocal (Mogg & Bradley, 2005; Peckham, McHugh & Otto, 2010). However, emotional Stroop paradigms have typically used emotional words as stimuli that only have a symbolic value of mood or threat, compared to more ecologically valid emotional stimuli such as images or facial expressions (Bradley et al. 1997; van Honk et al. 2001).

There has been little research on emotional Stroop performance for those classed at risk of mental health problems and to our knowledge nothing on children/adolescents. Though, Chung and Jeglic (2017), showed that a group of young adults at risk of suicide had longer latencies to suicide-related words that was significant predictor of suicide risk.

We have here a community sample of adolescents aged 11-18. That have been assessed on measures of anxiety and depression. Around a quarter of the sample meet the threshold for risk of depression. We developed a version of the EST that requires identifying happy and sad emotional words while inhibiting facial expressions of pictures of adolescents.

We hypothesise that:

1. The EST will show the typical Stroop effect with participants responding quickest to congruent stimuli, then neutral stimuli and slowest to incongruent stimuli
2. There will be an emotional Stroop effect with longer latencies for negative (i.e. sad) stimuli.
3. There will be an age effect with older participants responding faster and more accurately than younger participants.
4. Those at risk of depression, and higher in ratings of anxiety, will be more susceptible to negative stimuli than those not at risk.

Method

Participants

Four hundred and eighty-five participants (66.98% females) aged 11-18 years (M age 14.4, SD 1.8) were recruited from schools in Greater London and Cambridge. All students aged 11-18 years from these schools were given an information sheet with details about the study and an opt-in consent form. This consent form was to take home to a parent or caregiver for those aged under 16 years. After signed completed consent forms were returned, participants were selected from each school or college on a first-come first-served basis.

The study was approved by the University of Cambridge and UCL Research Ethics Committees. Participants were given a £15 shopping vouchers for taking part in testing.

Measures

The Emotional Stroop Task (EST)

The EST used here was adapted from one used by Preston and Stansfield (2008) and was programmed in E-Prime version 2.0 (Schneider, Eschman, & Zuccolotto, 2002). The adaptations to the previous version were the replacing of adult faces with adolescent faces, changing the affective words used to more age-appropriate ones and simplifying the task so that there were only two response options (happy or sad) rather than three (happy, sad or angry).

Stimuli comprised composite pictures of faces with words. We used 12 pictures of faces (three different facial expression from four different adolescent actors) taken from the National Institute of Mental Health Child Emotional Faces Picture Set (NIMH-ChEFS; Egger, et al., 2011). Each actor displayed a face with a happy expression, a sad expression, or the face was neutral. For the neutral condition, rather than present a face with a neutral expression, the decision was taken to scramble the image as prior research suggests that a neutral expression is perceived as more similar to a sad expression than a happy one (e.g. Thomas, et al., 2001).

Each face then had a different happy (Cheerful, Glad, Jolly, Joyful) or sad (Gloomy, Upset, Miserable, Hopeless) word superimposed semi transparently over it, centred vertically on the nose. The two word sets were matched in terms of age of acquisition (Kuperman, Stadthagen-Gonzalez & Brysbaert, 2012). We created 8 versions of the 12 faces (each face superimposed with each emotional word), giving 96 different face-word composite pictures in total. Each picture was presented twice giving a total of 192 trials; 64 of these were Congruent trials where the word was superimposed over a face with an expression that matched that word in terms of valence (e.g. a happy word superimposed over a happy expression), 64 were neutral trials where a word was superimposed over a scrambled face (e.g. a happy word superimposed over a neutral face), and 64 were incongruent trials where the word was superimposed over a face with an expression that was the opposite to the valence of the word (e.g. a happy word superimposed over a sad face). Figure 1 shows examples of each condition.

Participants were instructed to respond to the word by indicating if it was a happy or sad word and to ignore the facial expression. They were asked to respond "as quickly and as accurately as possible" by pressing a button on the computer keyboard labelled with an 'S' (for sad) or an 'H' (for happy).

A number of measures of performance can be calculated from the EST. Reaction time (RT) is the time taken in milliseconds (ms) for a correct response. Accuracy is the percentage of correct responses and is calculated by taking the total number of correct responses, dividing by the total number of trials and then multiplying by 100. Separate measures of RT, and accuracy were calculated for Congruent, Neutral and Incongruent conditions. Because the EST has an affective element, separate measures of RT, and accuracy were also calculated for Happy and Sad target conditions (i.e. whether the target word was a 'happy' word or a 'sad' word). Likewise, measures of RT, and accuracy were calculated for Happy, Neutral and Sad background faces (collapsed across congruent and incongruent conditions).The Stroop effect was calculated by subtracting performance on incongruent trials from performance on congruent trials. Separate Stroop effect indices were calculated for Happy and Sad conditions and affective indices were calculated by subtracting performance on Sad conditions from performance on Happy conditions.

Figure 1. Examples of Congruent, Neutral and Incongruent Stimuli from the Emotional Stroop Task

Center for Epidemiologic Studies Depression Scale (CES-D)

The CES-D (Radloff, 1977) is a 20 item questionnaire in which the participant rates how often over the past week they have experienced symptoms associated with depression. Responses range from 0 to 3 with 0 = rarely or none of the time, 1 = some or little of the time, 2 = moderately or much of the time and 3 = most or almost of the time. Responses are summed with higher scores indicating greater depressive symptoms. Although different at risk cut-offs have been previously used for the CES-D, a score over 19 was chosen based here based on data from a large (N=5262) school-based epidemiological study (Briere, Pascal, Dupere & Janosz, 2013).

Cattell Culture Fair Intelligence Test (CFIT)

The CFIT (Cattell, 1963) is a measure of non-verbal intelligence devoid of sociocultural and environmental influences. A paper and pencil version of Scale 2 Form A of the test was used. This comprises four timed subtests that consist of questions involving the relationships between pictures of abstract geometric shapes (e.g. choosing a shape that is in some way different from others). Correct responses are summed and age-appropriate standard scores are calculated based on a set of existing norms.

Revised Children's Anxiety and Depression Scale (RCADS)

The RCADS (Chorpita, Yim, Moffitt, Umemoto & Francis, 2000) is a 15-item self-report questionnaire that measures symptoms of anxiety and low mood. Participants record on a four-point scale the frequency of their symptoms (0= Never; 1= Sometimes; 2=Often; 3=Always). The items in the RCADS are summed to give a total score ranging from 0-40 with greater scores indicating greater symptoms of anxiety.

Procedure

Participants were tested in groups of between five and 12 and supervised by at least two researchers. Before starting the EST, a researcher added stickers that read 'S' and 'H' to the buttons '1' and '2' on the keyboard. The order in which the buttons appeared was by randomised across participants. Before starting the Emotional Stroop task, participants were given a set of instructions (see Supplementary Material) and completed a series practice trials. When the participants finished the task they completed the CES-D, the RCADS and CFIT.

Results

Participants

Data from 18 participants was removed from analysis because either: they scored below 70 standard score points on the IQ test (n=1), they did not complete the Emotional Stroop Task (n=7) or failed to get more than 75% of trials correct (n=10). In addition, four participants were missing IQ data, 12 did not complete the CES-D and 16 did not complete the RCADS. These participants were removed from the analysis of these measures. Prior to analysis, individual RTs on the EST that were more than three SDs from each individual's mean RT were identified as outliers and set aside (see Preston & Stansfield, 2008), resulting in an average data loss of 8.26% per participant.

Emotional Stroop Performance

Table 1 shows the mean, standard deviation and range of scores for the main variables included in the study.

Table 1.

Means (sd) and range of the main variables in the study

	Mean (sd)	Range
Age (months)	14.4 (1.80)	11.2 - 18.5
Emotional Stroop Task		
Congruent RT (ms)	858.77 (234.08)	480.80 - 1978.87
Neutral RT (ms)	896.54 (253.08)	459.85 - 2444.30
Incongruent RT (ms)	918.15 (264.30)	456.65 - 2626.80
Congruent Accuracy (%)	95.28 (4.23)	76.67 - 100
Neutral Accuracy (%)	94.32 (5.01)	71.43 - 100
Incongruent Accuracy (%)	91.4 (6.18)	66.13 - 100
Stroop valance		
Words		
Happy RT (ms)	875.8 (242.46)	450.27 - 2101.65
Sad RT (ms)	893.23 (246.12)	469.25 - 2210.78
Happy Accuracy (%)	93.99 (4.43)	76.48 - 100
Sad Accuracy (%)	93.5 (4.90)	72.38 - 100
Faces		
Happy RT (ms)	873.9 (244.63)	464.31 - 2296.01
Neutral RT (ms)	890.82 (251.22)	457.57 - 2349.55
Sad RT (ms)	888.82 (246.22)	457.40 - 2242.15
Neutral Accuracy (%)	3.32 (1.27)	.95 - 9.21
Happy Accuracy (%)	93.14 (5.24)	71.22 - 100
Neutral Accuracy (%)	94.4 (4.94)	72.81 - 100
Sad Accuracy (%)	93.68 (5.21)	75.81 - 100
IQ (CFIT standard scores)	113.2 (17.07)	76 - 158
Depression (CES-D total)	17.38 (10.69)	0 - 56
Anxiety (RCADS Total)	13.04 (8.11)	0 - 40

Note: IQ = Cattell Culture Fair Intelligence Test; CES-D = Center for Epidemiological Studies-Depression Scale; RCADS = Revised Children's Anxiety and Depression Scale;

ANOVAs to explore within-participant differences in RT between Congruent, Neutral and Incongruent conditions of the EST showed that, as hypothesised, participants responded significantly faster to Congruent than Neutral stimuli ($F(1, 466) = 91.25$, $p < .001$, $d = .16$), to Congruent than Incongruent stimuli ($F(1,446) = 224.29$, $p < .001$, $d = .24$) and to Neutral than Incongruent stimuli ($F(1,446) = 25.73$, $p < .001$, $d = .08$). Participants also made significantly fewer errors in response to

Congruent than Neutral stimuli (F (1,466) =24.09, p<.001, d=.21), Congruent than Incongruent stimuli (F (1,466) =236.97, p<.001, d=.73) and Neutral than Incongruent stimuli (F (1,466) =179.29, p<.001, d=.52). Figure 2 shows RTs for the EST for Congruent, Neutral and Incongruent conditions. Figure 3 shows accuracy rates for Congruent, Neutral and Incongruent conditions.

Further ANOVAs to examine the within-participants effects of valence showed that participants had faster RTs to Happy than Sad words (F (1,466) = 37.80, p=<.001, d=.07),and had a greater degree of accuracy for Happy than Sad words (F (1,466) = 8.77, p=.003, d=.11). Participants also responded faster to Happy faces than Sad faces (F (1,466) = 15.43, p=<.001, d=.06) or Neutral faces (F (1,466) = 16.44, p=<.001, d=.07). Participants showed a greater degree of accuracy for Neutral than Sad (F (1,466) = 4.725, p=.030, d=.25) or Happy faces (F (1,466) = 11.362, p=.001, d=.14). There were no significant differences in accuracy between Happy faces and Sad faces and there was no difference in RT or accuracy between the faces and words conditions. (all βs<-.07 all ps>.17). Figure 4 shows within participant performance for affective condition for both words and Figure 5 shows within participant performance for faces.

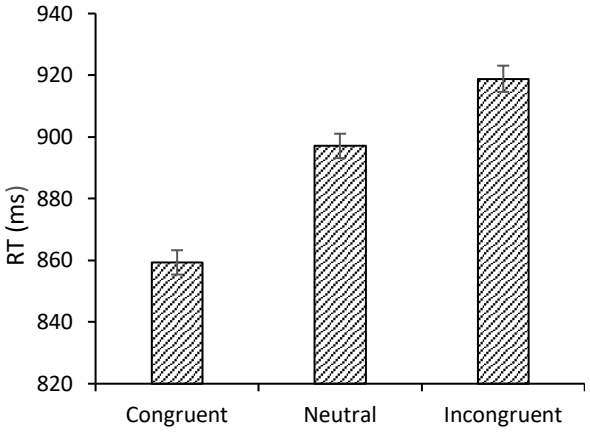

Figure 2. Mean Reaction Times for each condition

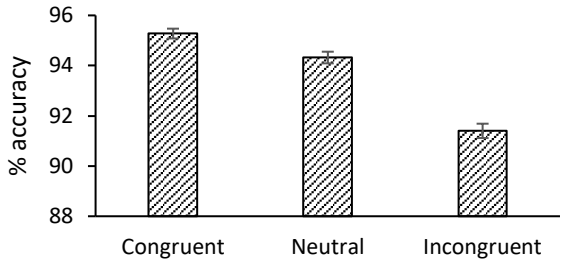

Figure 3. Mean Accuracy rate for each condition

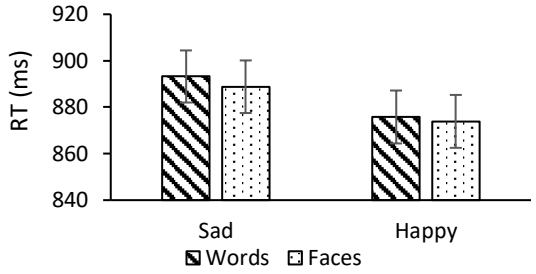

Figure 4. Mean Reaction Time by affective condition

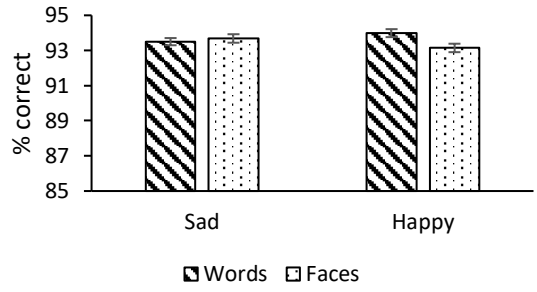

Figure 5. Mean Accuracy rate by affective condition

Separate Stroop indices (incongruent condition minus congruent condition RT) were calculated for both happy and sad conditions. For example, the Happy Stroop Index would be the RT

for a happy word over a happy face minus the RT for a happy word over a sad face. ANOVAs for each showed that there was a significant difference in RT between the Sad and Happy Stroop indices ($F(1,466) = 15.428$, $p<.001$, $d=.26$), with the Stroop effect greater for happy stimuli (Figure 6).

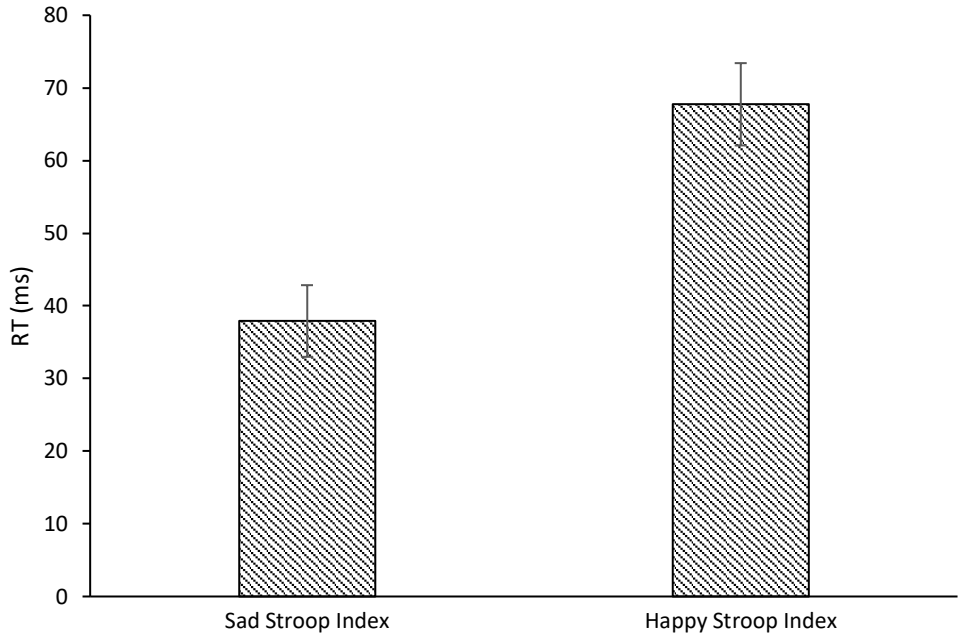

Figure 6. Sad and Happy Stroop Indices

Emotional Stroop performance by age, IQ and gender

Linear regressions were used to examine if performance on the EST improved with age or level of non-verbal intelligence, and to examine the relationship between EST performance and gender. Zero-order correlations (Table 2) revealed some significant associations between these independent variables as well as with measures of depression and anxiety. Consequently the regression models adjusted for these potential confounds in the first block. The variable of interest (age, IQ or gender) was entered in the second block and each Stroop measure (RT, accuracy) entered as the dependent variable.

Table 2.

Inter-correlations between independent variables

	Gender	IQ	Depression	Anxiety
Age	-.18**	-.01	.17**	.01
Gender		.10*	-.25**	-.23**
IQ			.04	.04
Depression				.64**

NB: IQ= CCFIT; Depression =CES-D; Anxiety=RCADS

For age (covarying gender, depression and anxiety), a significant regression equation was found in RTs for Congruent ($R^2_{adjusted}$ =.156, F (4, 435) =21.069, p<.001, β= -.408, p<.001), Neutral ($R^2_{adjusted}$ =.146, F (4, 435) =19.549, p<.001, β= -.393, p<.001) and Incongruent stimuli ($R^2_{adjusted}$ =.150, F (4, 435) =20.149, p<.001, β= -.403, p<.001) with older participants faster than younger participants in all cases. There was no significant age-related difference in accuracy or for the Stroop effect ($R^2_{adjusted}$ =.003, F (4, 435) =1.313, p=.264, β= -.408, p<.001) with a bigger Stroop effect shown in younger participants.

A similar pattern of results was shown for IQ (covarying for gender) with significant regression equations found for RTs for Congruent ($R^2_{adjusted}$ =.035, F (2, 463) =9.508, p<.001, β= -.196, p<.001), Neutral ($R^2_{adjusted}$ =.040, F (2, 463) =10.701, p<.001, β= -.203, p<.001) and Incongruent stimuli ($R^2_{adjusted}$ =.030, F (2, 463) =8.147, p<.001, β= -.184, p<.001), with those with higher IQs reacting faster than those with lower IQs in all cases. Finally, there were significant IQ-related differences in accuracy for Congruent stimuli ($R^2_{adjusted}$ =.030, F (2, 463) =8.237, p<.001, β= .102, p=.027), with those with higher IQs making fewer errors. There was no significant result for Neutral and Incongruent accuracy or for the Stroop effect (all βs <.048 all ps>.288).

For gender (covarying IQ, Depression and Anxiety), results showed that females had faster RTs in response to congruent stimuli ($R^2_{adjusted}$ =.047, F (4, 439) =6.367, p<.001, β= .101, p=.039) and neutral stimuli ($R^2_{adjusted}$ =.050, F (4, 439) =6.764, p<.001, β= .112, p=.022), but not incongruent

stimuli (β=.091, p=.063). Females made significantly fewer errors than males for congruent ($R^2_{adjusted}$ =.031, F (4, 439) =4.532, p=.001, β=-.175, p<.001), neutral ($R^2_{adjusted}$ =.044, F (4, 439) =6.065, p<.001, β=-.225, p<.001) and incongruent ($R^2_{adjusted}$ =.029, F (4, 439) =4.249, p=.002, β=-.199, p<.001) stimuli. There were no gender differences for the Stroop effect (all βs <.048 all ps>.288).

Similar regressions, but this time with the difference between Happy and Sad conditions as the dependent variable, were ran to explore the effects of age/IQ/gender on task valance. All results were non-significant (all ps>.08).

Emotional Stroop performance and mental health

For analysis of depression, participants were allocated to two groups based on CES-D scores, with scores of 19 and over indicating a risk of depression and scores less than 19 indicating no risk (see Briere, et al., 2013).

Linear regressions were used to assess if there were any differences on the main EST measures (RT, accuracy, Stroop effect). All results were non-significant (all p>.07).

To explore if there were group differences for valence on the EST as a function of risk of depression, a series of mixed-model general linear models were conducted with EST condition (happy, sad) as the within-subjects factor, Group ('at risk', 'not at risk') as the between-subjects factor and age and gender included as covariates. Separate models were constructed for each EST variable (Stroop, RT, accuracy). There were no other significant interactions (all Fs<1, all ps>.5)[1].

To examine the effect of anxiety ratings (RCADS) on Stroop performance, a series of multiple linear regressions were carried out. Because of the correlation between RCADS total and gender, all models included gender as a covariates in block 1, RCADS total was included in block 2, and

[1] Analysis of CED-D using total score rather than subgroups, showed the same pattern of results.

individual emotional Stroop measures (RT, accuracy) included as separate dependent variables. The results of the regression showed no significant differences between any conditions for were explained by any RCADS total.

For valence, a series of similar linear regressions were conducted, the only alteration was with the difference between Happy and Sad conditions was entered as the dependent variable. These showed that no significant variance was explained by scores of anxiety (all ps>.08)

References

Bradley, B. P., Mogg, K., Millar, N., Bonham-Carter, C., Fergusson, E., Jenkins, J., & Parr, M. (1997). Attentional biases for emotional faces. *Cognition & Emotion*, *11*(1), 25-42.

Bremner, J. D., Vermetten, E., Vythilingam, M., Afzal, N., Schmahl, C., Elzinga, B., & Charney, D. S. (2004). Neural correlates of the classic color and emotional stroop in women with abuse-related posttraumatic stress disorder. *Biological psychiatry*, *55*(6), 612-620.

Brière, F. N., Pascal, S., Dupéré, V., & Janosz, M. (2013). School environment and adolescent depressive symptoms: a multilevel longitudinal study. *Pediatrics*, *131*(3), e702-e708.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of educational psychology*, *54*(1), 1.

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour research and therapy*, *38*(8), 835-855.

Chung, Y., & Jeglic, E. L. (2017). Detecting suicide risk among college students: A test of the predictive validity of the modified emotional Stroop task. *Suicide and Life-Threatening Behavior*, *47*(4), 398-409.

Egger, H. L., Pine, D. S., Nelson, E., Leibenluft, E., Ernst, M., Towbin, K. E., & Angold, A. (2011). The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): a new set of children's facial emotion stimuli. *International journal of methods in psychiatric research*, *20*(3), 145-156.

Gotlib, I. H., & McCann, C. D. (1984). Construct accessibility and depression: An examination of cognitive and affective factors. *Journal of personality and social psychology*, *47*(2), 427.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior research methods*, *44*(4), 978-990.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin*, *109*(2), 163.

Mogg, K., & Bradley, B. P. (2005). Attentional bias in generalized anxiety disorder versus depressive disorder. *Cognitive therapy and research*, *29*(1), 29-45.

Peckham, A. D., McHugh, R. K., & Otto, M. W. (2010). A meta-analysis of the magnitude of biased attention in depression. *Depression and anxiety*, *27*(12), 1135-1142.

Preston, S. D., & Stansfield, R. B. (2008). I know how you feel: Task-irrelevant facial expressions are spontaneously processed at a semantic level. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(1), 54-64.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, *1*(3), 385-401.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's Guide. Reference Guide. Getting Started Guide*. Psychology Software Tools, Incorporated.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Thomas, K. M., Drevets, W. C., Dahl, R. E., Ryan, N. D., Birmaher, B., Eccard, C. H., ... & Casey, B. J. (2001). Amygdala response to fearful faces in anxious and depressed children. *Archives of general psychiatry*, *58*(11), 1057-1063.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological bulletin*, *120*(1), 3.