Jamieson and Harkins (2007)

EPPI-Centre (2003) & Critical Appraisal Skills Programme (2018)

If the study has a broad focus and this data extraction focuses on just one component of the study, please specify this here

☒ Not applicable (whole study is focus of data extraction)

☐ Specific focus of this data extraction (please specify)

Study aim(s) and rationale

Was the study informed by, or linked to, an existing body of empirical and/or theoretical research?

Please write in authors' declaration if there is one. Elaborate if necessary, but indicate which aspects are reviewers' interpretation.

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

- Stereotype threat
- Working memory interference account vs mere effort account

Do authors report how the study was funded?

☐ Explicitly stated (please specify)

☒ Implicit (please specify)

☐ Not stated/unclear (please specify)

Special thanks to Neal Pearlmutter for all his help and advice in the conduct of this research and for use of the eye-tracker (supported by National Institutes of Health Grant R01-DC05237). Thanks also go to Sean Allen for writing the computer programs used in this work, Richard Gramzow for his helpful comments during the writing of this paper, and Rhea Eskew for his assistance with data analyses.

Study research question(s) and its policy or practice focus

What is/are the topic focus/foci of the study?

- Does the theory of the mere effort account, account better for the effects of stereotype threat and can it better predict changes than the working memory interference account?

What is/are the population focus/foci of the study?

- not stated

What is the relevant age group?

☐ Not applicate (focus not learners)

☐ 0 - 4

☐ 5 - 10

☐ 11 - 16

☐ 17 - 20

☐ 21 and over

☒ Not stated/unclear

What is the sex of the population focus/foci?

☐ Not applicate (focus not learners)

☐ Female only

☐ Male only

☐ Mixed sex

☒ Not stated/unclear

What is/are the educational setting(s) of the study?

☐ Community centre

☐ Correctional institution

☐ Government department

☐ Higher education institution

☐ Home

☐ Independent school

☐ Local education authority

☐ Nursery school

☐ Other early years setting

☐ Post-compulsory education institution

☐ Primary school

☐ Residential school

☐ Secondary school

☐ Special needs school

☐ Workplace

☐ Other educational setting

In Which country or cuntries was the study carried out?

☒ Explicitly stated (please specify)

☐ Not stated/unclear (please specify)

- United states

Please describe in more detail the specific phenomena, factors, services, or interventions with which the study is concerned

What are the study reserach questions and/or hypotheses?

Research questions or hypotheses operationalise the aims of the study. Please write in authors' description if there is one. Elaborate if necessary, but indicate which aspects are reviewers' interpretation.

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

Experiment 1: Because the working memory inference account predicts that the experience of threat will reduce working memory capacity, participants subject to threat should perform more poorly than controls regardless of target display time. Specifically, working memory resources are required to inhibit the prepotent response to look towards the cue and to generate saccades to the target. If working memory is impaired, then participants should look towards the cue more often and take longer to launch saccades to the target than control participants.
The mere effort account also predicts that, at some display time, participants in the stereotype threat condition will perform more poorly than participants in the control condition because the participants will not have sufficient time to correct for the prepotent response, looking at the cue. However, at some longer display time, there will be no difference, and at a longer display time, a reversal, as the threat participants do have time to make the correction.

Experiment 2: At this exposure time, the predictions of the working memory interference and mere effort accounts diverge. Mere effort predicts that participants subject to threat will report target orientation either as quickly as or more quickly than control participants because the additional time should provide them with the opportunity to correct for the potentiated prepotent response and still see the target, and they are motivated to do so. If the experience of stereotype threat impairs working memory capacity, then stigmatized individuals should perform more poorly than the control group on the antisaccade task regardless of display time.

Experiment 3: Examine the specific processes that produce the performance outcomes observed in Experiments 1 and 2, by using eye-tracking technology to measure participants' eye movements during the antisaccade task.

Experiment 4: We predicted that under high cognitive load, stereotype threat participants should no longer have the working memory capacity necessary to produce their performance advantage. The potentiated tendency to look at the cue produced by stereotype threat in combination with the reduced ability to inhibit looking at the cue produced by the high-load concurrent task should increase the likelihood that participants in the stereotype threat condition will produce reflexive sacades. In addition, reduced working memory capacity should negatively impat the advantage of the advantage in saccade launch latency enjoyed by threat participants. Under low-cognitive load, however, the pattern of performance should replicate the pattern observed in Experiments 2 and 3: Participants in stereotype threat condition should have faster reaction times than participants in the control group.

Methods - Design

Which variables or concepts, if any, does the study aim to measure or examine?

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

Experiment 1: - response time (RT) - antisaccade task performance - prosaccade task performance - stereotype threat manipulation - questionnaires: stereotype threat manipulation check, interest, anxiety, how they think they performed, effort they put in

Experiment 2: - same as Experiment 1

Experiment 3: - see Experiment 2 - eye-tracking

Experiment 4: - see Experiment 2 - n-back task

Study timing

Please indicate all that apply and give further details where possible.

If the study examines one or more samples, but each at only one point in time it is cross-sectional.
If the study examines the same samples, but as they have changed over time, it is retrospective, provided that the interest is in starting at one timepoint and looking backwards over time.
If the study examines the same samples as they have changed over time and if data are collected forward over time, it is prospective provided that the interest is in starting at one timepoint and looking forward in time.

☒ Cross-sectional

☐ Retrospective

☐ Prospective

☐ Not stated/unclear (please specify)

If the study is an evaluation, when were measurements of the variable(s) used for outcome made, in relation to the intervention?

If at least one of the outcome variables is measured both before and after the intervention, please use the before and after category.

☐ Not applicable (not an evaluation)

☒ Before and after

☐ Only after

☐ Other (please specify)

☐ Not stated/unclear (please specify)

Methods - Groups

If comparisons are being made between two or more groups, please specify the basis of any divisions made for making these comparisons.

Please give further details where possible.

☐ Not applicable (not more than one group)

☒ Prospecitive allocation into more than one group (e.g. allocation to different interventions, or allocation to intervention and control groups)

☐ No prospective allocation but use of pre-existing differences to create comparison groups (e.g. receiving different interventions, or characterised by different levels of a variable such as social class)

☐ Other (please specify)

☐ Not stated/unclear (please specify)

How do the groups differ?

☐ Not applicable (not more than one group)

☒ Explicityly stated (please specify)

☐ Implicit (please specify)

☐ Not stated/unclear (please specify)

Experiment 1: - Stereotype threat manipulation vs no stereotype threat manipulation.

Experiment 2: - Stereotype threat manipulation vs no stereotype threat manipulation.

Experiment 3: - Stereotype threat manipulation vs no stereotype threat manipulation.

Experiment 4: - n-back condition (0 back vs 2 back) - stereotype threat manipulation vs no stereotype threat manipulation.

Number of groups

For instance, in studies in which comparisons are made between groups, this may be the number of groups into which the dataset is divided for analysis (e.g. social class, or form size), or the number of groups allocated to, or receiving, an intervention.

☐ Not applicable (not more than one group)

☐ One

☒ Two

☐ Three

☐ Four or more (please specify)

☐ Other/unclear (please specify)

- two

Experiment 4: - four (2 [stereotype threat vs no threat] x 2 [0-back vs 2-back])

Was the assignment of participants to interventions randomised?

☐ Not applicable (not more than one group)

☐ Not applicate (no prospective allocation)

☐ Random

☐ Quasi-random

☐ Non-random

☒ Not stated/unclear (please specify)

Where there was prospective allocation to more than one group, was the allocation sequence concealed from participants and those enrolling them until after enrolment?

Bias can be introduced, consciously or otherwise, if the allocation of pupils or classes or schools to a programme or intervention is made in the knowledge of key characteristics of those allocated. For example: children with more serious reading difficulty might be seen as in greater need and might be more likely to be allocated to the 'new' programme, or the opposite might happen. Either would introduce bias.

☐ Not applicable (not more than one group)

☐ Not applicable (no prospective allocation)

☒ Yes (please specify)

☐ No (please specify)

☐ Not stated/unclear (please specify)

Apart from the experimental intervention, did each study group receive the same level of care (that is, were they treated equally)?

☒ Yes
☐ No
☐ Can't tell

Study design summary

In addition to answering the questions in this section, describe the study design in your own words. You may want to draw upon and elaborate the answers you have already given.

Experiment 1: - target display time was 150 ms 1. consent 2. verbal and written instructions 3. practice trials 4. stereotype threat manipulation 5. antisaccade task or prosaccade task (random for each participant) 6. questionnaire for stereotype threat manipulation check 7. other task, either antisaccade or prosaccade, depending on which was done in the first round 8. questionnaire for stereotype threat manipulation check 9. questionnaire for interest, anxiety, how they think they performed, effort they put in

Experiment 2: - target display time was 250 ms - same as experiment 1, but participants were just given one questionnaire (manipulation check) upon completion of the first block of trails.

Experiment 3: - Manipulation check and questionnaire items were identical to those in Experiment 2 - target display time was 250 ms - Calibration test for eye-tracking was done during the experiment, an experimenter was informed when the calibration test screen appeared
- Questionnaire was filled out after the first saccade task.

Experiment 4: 1. six practice trails to familiarize participants with the n-back task and the saccade task 2. stereotype threat manipulation - Manipulations and questionnaires were the same as in Experiment 2 - Participants were presented a letter displayed in the centre of the screen, followed by a saccade trail. After the saccade trail, another letter was then presented, followed by another saccade trail. This pattern of letter-then-saccade trail was repeated throughout the block of trails. The participants was to indicate whether the current letter matched the letter that appeared two places back in the sequence and to perform well on the saccade trails. In the control condition (0-back task), participants were still required to remember two letters, the two letter remained the same throughout the task
- additional question in the questionnaire, asked participants to assess the difficulty of the n-back task.

Methods - Sampling strategy

Are the authors trying to produce findings that are representative of a given population?

Please write in authors' description. If authors do not specify please indicate reviewers' interpretation.

☐ Explicitly stated (please specify)
☐ Implicit (please specify)
☒ Not stated/unclear (please specify)

Which methods does the study use to identify people or groups of people to sample from and what is the sampling frame?

e.g. telephone directory, electoral register, postcode, school listing, etc. There may be two stages – e.g. first sampling schools and then classes or pupils within them.

☐ Not applicable (please specify)
☒ Explicitly stated (please specify)
☐ Implicit (please specify)
☐ Not stated/unclear (please specify)

Experiment 1: - university course requirement

Experiment 2: - university course requirement

Experiment 3: - class credit

Experiment 4: - university course requirement

Which methods does the study use to select people or groups of people (from the sampling frame)?

e.g. selecting people at random, systematically - selecting for example every 5th person, purposively in order to reach a quota for a given characteristic.

☐ Not applicable (no sampling frame)
☒ Explicitly stated (please specify)
☐ Implicit (please specify)
☐ Not stated/unclear (please specify)

Experiment 1: - Normal or corrected-to-normal vision

Experiment 2: - Normal or corrected-to-normal vision

Experiment 3: - Normal or corrected-to-normal vision but none wore eye-glsses, which interfered with eye tracking

Experiment 4: - Normal or corrected-to-normal vision

Planned sample size

If more than one group please give details for each group separately.

☐ Not applicable (please specify)
☐ Explicitly stated (please specify)
☒ Not stated/unclear (please specify)

Methods - Recruitment and consent

Which methods are used to recruit people into the study?

e.g. letters of invitation, telephone contact, face-to-face contact.

☐ Not applicable (please specify)
☒ Explicitly stated (please specify)
☐ Implicit (please specify)
☐ Not stated/unclear (please specify)

Experiment 1 & 2 & 4: - University course requirement

Experiment 3: - Class credit

Were any incentives provided to recruit people into the study?

☐ Not applicable (please specify)
☒ Explicitly stated (please specify)
☐ Not stated/unclear (please specify)

Experiment 1 & 2 & 4: - participated in exchange for partial fulfilment of a course requirement

Experiment 3: - participated in exchange for class credit

Was consent sought?

Please comment on the quality of consent if relevant.

☐ Not applicable (please specify)
☒ Participant consent sought
☐ Parental consent sought
☐ Other consent sought
☐ Consent not sought
☐ Not stated/unclear (please specify)

Are there any other details relevant to recruitment and consent?

☒ No
☐ Yes (please specify)

Methods - Actual sample

What was the total number of participants in the study (the actual sample)?

If more than one group is being compared please give numbers for each group.

☐ Not applicable (e.g. study of policies, documents, etc)
☒ Explicitly stated (please specify)
☐ Implicit (please specify)
☐ Not stated/unclear (please specify)

Experiment 1: - Eighty Northeastern University undergraduate students (40 men and 40 women) participated in this experiment in exchange for partial fulfilment of a course requirement.

Experiment 2: - Thirty six female Northeastern University undergraduate students participated in this experiment in partial fulfilment of a course requirement.

Experiment 3: - Thirty six Northeastern University students participated in this experiment in exchange for class credit.

Experiment 4: - Seventy-two Northeastern University female undergraduates participated in this experiment in exchange for partial fulfilment of a course requirement.

What is the proportion of those selected for the study who actually participated in the study?

Please specify numbers and percentages if possible.

☐ Not applicable (e.g. study of policies, documents, etc)
☐ Explicitly stated (please specify)
☐ Implicit (please specify)
☒ Not stated/unclear (please specify)

Which country/countries are the individuals in the actual sample from?

If UK, please distinguish between England, Scotland, N. Ireland, and Wales if possible. If from different countries, please give numbers for each. If more than one group is being compared, please describe for each group.

☐ Not applicable (e.g. study of policies, documents, etc)

☐ Explicitly stated (please specify)

☒ Implicit (please specify)

☐ Not stated/unclear (please specify)

- Northeastern University undergraduate students (United States)

What ages are covered by the actual sample?

Please give the numbers of the sample that fall within each of the given categories. If necessary, refer to a page number in the report (e.g. for a useful table). If more than one group is being compared, please describe for each group. If follow-up study, age at entry to the study.

☐ Not applicable (e.g. study of policies, documents, etc)
☐ 0 to 4
☐ 5 to 10
☐ 11 to 16
☐ 17 to 20
☐ 21 and over
☒ Not stated/unclear (please specify)

What is the socio-economic status of the individuals within the actual sample?

If more than one group is being compared, please describe for each group.

☐ Not applicable (e.g. study of policies, documents, etc)
☐ Explicitly stated (please specify)
☐ Implicit (please specify)
☒ Not stated/unclear (please specify)

What is the ethnicity of the individuals within the actual sample?

If more than one group is being compared, please describe for each group.

☐ Not applicable (e.g. study of policies, documents, etc)
☐ Explicitly stated (please specify)
☐ Implicit (please specify)
☒ Not stated/unclear (please specify)

What is known about the special educational needs of individuals within the actual sample?

e.g. specific learning, physical, emotional, behavioural, intellectual difficulties.

☐ Not applicable (e.g. study of policies, documents, etc)
☐ Explicitly stated (please specify)
☐ Implicit (please specify)
☒ Not stated/unclear (please specify)

Is there any other useful information about the study participants?

☐ Not applicable (e.g. study of policies, documents, etc)
☒ Explicitly stated (please specify no/s.)
☐ Implicit (please specify)
☐ Not stated/unclear (please specify)

Experiment 1: - 40 men, 40 women

Experiment 2: - just women, no men

Experiment 3: - presumably just women again, same as in experiment 2 but this isn't mentioned explicitly

Experiment 4: - just women

How representative was the achieved sample (as recruited at the start of the study) in relation to the aims of the sampling frame?

Please specify basis for your decision.

☐ Not applicable (e.g. study of policies, documents, etc)
☒ Not applicable (no sampling frame)
☐ High (please specify)
☐ Medium (please specify)
☐ Low (please specify)
☐ Unclear (please specify)

If the study involves studying samples prospectively over time, what proportion of the sample dropped out over the course of the study?

If the study involves more than one group, please give drop-out rates for each group separately. If necessary, refer to a page number in the report (e.g. for a useful table).

☐ Not applicable (e.g. study of policies, documents, etc)
☒ Not applicable (not following samples prospectively over time)
☐ Explicitly stated (please specify)
☐ Implicit (please specify)
☐ Not stated/unclear

For studies that involve following samples prospectively over time, do the authors provide any information on whether and/or how those who dropped out of the study differ from those who remained in the study?

☐ Not applicable (e.g. study of policies, documents, etc)
☒ Not applicable (not following samples prospectively over time)
☐ Not applicable (no drop outs)
☐ Yes (please specify)
☐ No

If the study involves following samples prospectively over time, do authors provide baseline values of key variables such as those being used as outcomes and relevant socio-demographic variables?

☐ Not applicable (e.g. study of policies, documents, etc)
☒ Not applicable (not following samples prospectively over time)
☐ Yes (please specify)
☐ No

Methods - Data collection

Please describe the main types of data collected and specify if they were used (a) to define the sample; (b) to measure aspects of the sample as findings of the study?

☐ Details

Experiment 1 & 2: - task performance -> b - questionnaire -> b

Experiment 3: - same as experiment 1 and 2 - eye-tracking -> b

Experiment 4: - Same as experiment 2 - question about difficulty of n-back task -> b - n-back task -> b

Which methods were used to collect the data?

Please indicate all that apply and give further detail where possible.

☐ Curriculum-based assessment
☐ Focus group
☐ Group interview
☐ One to one interview (face to face or by phone)
☐ Observation
☐ Self-completion questionnaire
☐ Self-completion report or diary
☐ Exams
☐ Clinical test
☐ Practical test
☐ Psychological test
☐ Hypothetical scenario including vignettes
☐ School/college records (e.g. attendance records etc)
☐ Secondary data such as publicly available statistics
☐ Other documentation
☐ Not stated/unclear (please specify)

Details of data collection methods or tool(s).

Please provide details including names for all tools used to collect data and examples of any questions/items given. Also please state whether source is cited in the report.

☒ Explicitly stated (please specify)
☐ Implicit (please specify)
☐ Not stated/unclear (please specify)

Experiment 1 & 2: - antisaccade task - prosaccade task - stereotype threat manipulation (similar to previous research, e.g., Brown & Pinel 2003; Keller & Daunenheimer, 2003; O'Brien & Crandall, 2003; Spencer et al., 1999) - stereotype threat manipulation check (twice), "To what extent are there gender differences in performance on this task?" on a 11-point Likert scale and "Who do you believe performs ebtter on this task?" on a 11-point Likert scale. - Participants were also asked to rate the extent to which they felt

that they could evaluate their performance and the extent to which their performance could be evaluated by the experimenter. Finally, they were asked to rate how interesting the task was, how anxious they felt about their performance, how well they thought they performed, and how much effort they put into the task, all on 11-point scales.

Experiment 3: - same as experiment 1 and 2 - eye-tracking data were collected using a Dr. Bouis infra-red oculometer (Dr. Bouis Devices, Karlsruhe, Germany) interfaced with the computer.

Experiment 4: - same as experiment 2 - n-back task - question about difficulty of n-back task

Who collected the data?

Please indicate all that apply and give further detail where possible.

☒ Researcher
☐ Head teacher/Senior management
☐ Teaching or other staff
☐ Parents
☐ Pupils/students
☐ Governors
☐ LEA/Government officials
☐ Other education practitioner
☐ Other (please specify)
☐ Not stated/unclear

Do the authors describe any ways they addressed the reliability of their data collection tools/methods?

e.g. test-retest methods (Where more than one tool was employed please provide details for each.)

☐ Details

Do the authors describe any ways they have addressed the validity of their data collection tools/methods?

e.g. mention previous validation of tools, published version of tools, involvement of target population in development of tools. (Where more than one tool was employed please provide details for each.)

☐ Details

Was there concealment of study allocation or other key factors from those carrying out measurement of outcome – if relevant?

Not applicable – e.g. analysis of existing data, qualitative study. No – e.g. assessment of reading progress for dyslexic pupils done by teacher who provided intervention. Yes – e.g. researcher assessing pupil knowledge of drugs - unaware of pupil allocation.

☐ Not applicable (please say why)
☐ Yes (please specify)
☐ No (please specify)

Where were the data collected?

e.g. school, home.

☒ Explicitly stated (please specify)

☐ Implicit (please specify)

☐ Unclear/not stated (please specify)

- the lab

Are there other important features of data collection?

e.g. use of video or audio tape; ethical issues such as confidentiality etc.

☐ Details

Methods - Data analysis

Which methods were used to analyse the data?

Please give details e.g. for in-depth interviews, how were the data handled? Details of statistical analysis can be given next.

☒ Explicitly stated (please specify)
☐ Implicit (please specify)
☐ Not stated/unclear (please specify)

Experiment 3: - Filters were used prior to data analysis to ensure that eye movements recorded by the eye tracker represented responses to the stimuli and were not random movements

Which statistical methods, if any, were used in the analysis?

☐ Details

Experiment 1 and 2: *Manipulation Check*: - 2 (condition) x 2 (gender) x 2 (task order: antisaccade first vs. prosaccade first) x 2 (task: antisaccade vs prosaccade) ANOVAs - Condition, gender, and task order were analysed as between-subjects effects, and task was analysed as a within-subjects effect.
Performance: - Performance data (accuracy and reaction time measured from the onset of the target) were analysed in 2 (condition) x 2 (gender) x 2 (task order) x 2 (task) ANOVAs. - Condition, gender, and task order were analysed as between-subjects effects, and task was analysed as a within-subjects effect.
Ancillary Measures: - Tukey's Honestly Significant Difference test (Kirk, 1995)

Experiment 3: - same as experiment 1 and 2 *Eye Movement Measures*: - Analyses were conducted on the three types of saccades produced on the antisaccade task—reflexive

saccades, corrective saccades, and correct saccades and on the eye movement data for the prosaccade task - Adjusted reaction time data for antisaccade trails were also analysed - The percentage and latency of reflexive saccades were analysed in 2 (condition) x 2 (task order) ANOVAs - The latencies of corrective saccades were analysed in a 2 (condition) x 2 (task order) x 2 (task) ANOVA - The latencies for correct saccades were analysed in a 2 (condition) x 2 (task order) x 2 (task) ANOVA - The latencies on the prosaccade tasks were analysed in a 2 (condition) x 2 (task order) ANOVA - The adjusted reaction times were analysed in a 2 (condition) x 2 (task order) x 2 (type of saccade) ANOVA

Experiment 4: *Manipulation Check*: - ST manipulation checks and other questionnaire items were analysed in 2 (stereotype threat condition) x 2 (saccade task) x 2 (n-back task) ANOVAs with condition, task, and n-back task as between-subjects factors
n-Back Task: - The percentrage of correct responses on the n-back task was analysed in a 2 (stereotype threat condition) x 2 (saccade task) x 2 (task order) x 2 (n-back task) ANOVA with condition, task order, and n-back task as between-subjects factors and saccade task as a within-subjects factor
Saccade tasks: - Accuracy and reaction time measured form target onset were analysed in 2 (ST condition) x 2 (saccade task) x 2 (task order) x 2(n-back task) ANOVAs, with condition, task order, and n-back task as between-subjects factors and saccade task as a within-subjects factor

What rationale do the authors give for the methods of analysis for the study?

e.g. for their methods of sampling, data collection, or analysis.

☐ Details

For evaluation studies that use prospective allocation, please specify the basis on which data analysis was carried out.

'Intention to intervene' means that data were analysed on the basis of the original number of participants as recruited into the different groups. 'Intervention received' means data were analysed on the basis of the number of participants actually receiving the intervention.

☐ Not applicable (not an evaluation study with prospective allocation)
☐ 'Intention to intervene'
☐ 'Intervention received'
☐ Not stated/unclear (please specify)

Do the authors describe any ways they have addressed the reliability of data analysis?

e.g. using more than one researcher to analyse data, looking for negative cases.

☐ Details

Do the authors describe any ways they have addressed the validity of data analysis?

e.g. internal or external consistency; checking results with participants.

☐ Details

Do the authors describe strategies used in the analysis to control for bias from confounding variables?

☐ Details

Please describe any other important features of the analysis.

☐ Details

Please comment on any other analytic or statistical issues if relevant.

☐ Details

Results and Conclusions

How are the results of the study presented?

e.g. as quotations/figures within text, in tables, appendices.

☐ Details

Experiment 1: - Figure - in text

Experiment 2: - in text

Experiment 3: - figures - tables - in text

Experiment 4: - in text

What are the results of the study as reported by authors?

Please give details and refer to page numbers in the report(s) of the study where necessary (e.g. for key tables).

☐ Details

Experiment 1: *Manipulation check*: -Participants in the stereotype threat condition reported that gender differences existed to a greater extent than participants in the control condition - Threatened participants reported that men performed these tasks better than women to a greater extent than participants in the control condition - These results indicate that the stereotype threat manipulation used in the present experiment was successful - Participants in the threat condition were aware of the negative group stereotype, and women were expected to perform more poorly than men.
Performance: Accuracy: - replicating past work, participants correctly reported the orientation of a greater percentage of targets in the prosaccade task than in the antisaccade task - This finding is expected because on the prosaccade task, unlike the antisaccade task, good performance does not require the inhibition of the prepotent response tendency. - Each of

the other reliable effects in the overall analysis, gender; Condition x Gender, and Task x Gender, must be interpreted in the context of the Condition x Task x Gender interaction. -On the antisaccade task, females in the stereotype threat condition performed more poorly than participants in any other condition, which did not differ from each other. - Thus, only the women subject to stereotype threat experienced performance decrements in their ability to accurately report target orientation. - There were no reliable differences among the condition on the prosaccade task - the largest difference between the pairs of these means was only 1.1%

Reaction time: - Replicating past work, we found a main effect of task - Participants responded to targets more quickly in the prosaccade task than in the antisaccade task - There was also a marginal gender main effect, with men tending to respond more quickly than women across tasks. - A Condition X Task Order interaction resulted from the fact that in the stereotype threat condition, it did not matter whether the antisaccade task or the prosaccade task came first. - In the control condition, participants completing the prosaccade task first responded more quickly across both tasks than participants who responded to the antisaccade task first - There was also a significant Condition x Task interaction - This interaction resulted from the fact that participants in both the stereotype threat and control conditions responded to the target reliably more quickly on prosaccade trails than on antisaccade trails, but the difference was greater in the stereotype threat condition. - As a result of differences in accuracy, the means of participants in the stereotype threat and control groups are based on different numbers of trails - Thus, any analysis comparing these groups on a measure of reaction time is suspect - In any event, these interactions were not predicted, nor do they alter the interpretation of results in the present experiment.

Ancillary Measures: - Analysis of the difficulty measure revealed a Gender x Task interaction - a THSD test showed that, on the prosaccade task, women and men did not differ in their ratings of task difficulty, but female participants in the stereotype threat condition rated the antisaccade task as more difficult than male participants. - The gender main effect must be interpreted in the context of this interactoin - Participants did not differ in their ratings of the extent the experimenter knew how well they performed or in their ratings of the extent to which they could evaluate their own performances - Analyses of the self-reports of anxiety experienced during task performance, task interest, how well they thought they performed, and effort were also all nonsignificant.

Experiment 2: *Manipulation Check*: - Participants in the stereotype threat condition reported that gender differences existed on the saccade tasks to a greater extent than participants in the control condition - The threatened participants also reported that men performed the tasks better than women to a greater extent than participants in the control condition

Performance: Accuracy: - As in Experiment 1, participants correctly reported the orientation of the target more accurately in the prosaccade task than in the antisaccade task - No other effects were reliable - In the present experiment, the target in the antisaccade task was more easily identified than in Experiment 1 - This increase in accuracy likely results from the increase in display time.

Reaction time: - As in Experiment 1, participants responded to target orientation more quickly in the prosaccade condition than in the antisaccade condition - More important,

there was a main effect for stereotype threat - Participants subject to stereotype threat responded more quickly than control participants - Separate contrasts show that this pattern characterized performance on both antisaccade and prosaccade trials - No other effects were reliable

Ancillary Measures: - Paritcipants rated the antisaccade task more difficult than the prosaccade task - There were no reliable differences in the participants' ratings of self-reported anxiety, effort, how well they performed, or the potential experimenter or self-evaluation.

Experiment 3: *Manipulation Check*: - Stereotype threat manipulation was successful

Performance Accuracy: - Participants were more accurate on prosaccade trials than on antisaccade trails - Participants did not differ in their ability to identify target orientation as a function of stereotype threat

Reaction time: - Participants responded to target orientation more quickly in the prosaccade task than in the antisaccade task - Participants subject to stereotype threat identified the target orientation more quickly than control participants - Separate contrasts showed that this pattern characterized performance on both antisaccade and prosaccade trials

Eye Movement Measures: Reflexive saccades: - Participants in the stereotype threat condition launched reflexive saccades on a greater percentage of the trails than control participants - There was also a tendency for stereotype threat participants to launch these saccades more quickly than participants in the control condition

Corrective saccades: - Participants subject to stereotype threat launched corrective saccades more quickly than control participants - Finding is consistent with the mere effort account but not an account relying solely on working memory inference

Correct saccades: - Participants subject to stereotype threat launched correct saccades more quickly than control participants - This finding is consistent with the mere effort account and not with a working memory inference account that relies solely on a reduction of working memory capacity during task performance.

Prosaccades: - This analysis revealed a tendency for participants in the stereotype threat dcondition to launch prosaccades more quickly

Adjusted reaction times: - The adjusted reaction times for participants subject to stereotype threat were significantly faster than those for participants in the control group - This finding is consistent with the mere effort account and not with a working memory inference alone. - This analysis also revealed a main effect for type of response - Adjusted reaction times for corrective saccades were longer than for correct saccades - Participants in the stereotype threat condition pressed the response key an average of 100 ms faster than participants in the control group - This response advantage cannot be attributed to eye movement because the combination of latency to launch and travel time brought the stereotype threat and control participants to the target area at the same time, over 200 ms before the target even appeared. - These findings suggest that participants saw the target at the same time, at the 400-ms mark - Thus it was the motivation to press the key to make the response that produced the reaction time difference between the condition on this prosaccade trails - This finding is also consistent with the notion that participants in the stereotype threat condition are motivated to perform well

Ancillary Measures: - Participants rated the antisaccade task as more difficult - There were

no reliable differences in the participants' ratings of self-reported anxiety, effort, how well they thought they performed, or the potential experimenter or self-evaluation.

Experiment 4: *Manipulation Check*: - Participants subject to ST indicated that gender differences existed to a greater extent than participants in the controls - ST participants also indicated that men performed better than women to a greater extent than controls did - ST manipulation was successful
n-back task: - Participants identified the target letter with higher accuracy in the 2-back condition - This effect was expected because the 2-back condition requires working memory resources to keep updating the last two letters, whereas the 0-back condition only requires participants to match the current letter with one of two preassigned target letters Participants did not differ in their ability to identify the target letter as a function of the stereotype threat manipulation - No other effects were reliable
Saccade Tasks:
Accuracy: - Participants responded to target orientation with higher accuracy on prosaccade trails than on antisaccade trails - This effect was consistent with findings form Experiments 1 through 3 - In the low-load 0 back condition, participants were marginally more accurate in identifying target orientation than participants in the high-load 2-back condition - No other effects were reliable - As in Experiments 2 and 3, stereotype threat did not affect the participants' ability to correctly identify target orientation
Reaction time: - Consistent with our previous experiments, participants identified target orientation significantly more quickly on prosaccade trails than on antisaccade trails - This analysis also revealed a significant N-Back Task x Condition x Saccade Task interaction - The 0-back condition represented a minimal memory load and should not have taxed the participants' working memory - Consistent with this notion, replicating Experiments 2 and 3, on the antisaccade task, participants in the ST condition identified target orientation more quickly than control participants - In the 2 back condition, the performance advantage enjoyed by participants in the ST condition was not only eliminated but also reversed - Participants in the control condition responded more quickly on antisaccade trails than participants under ST - This finding is consistent with the prediction that introducing a taxing concurrent task will undermine the performance of ST participants - Replicating Experiments 2 and 3, on the prosaccade task, ST participants in the 0-back condition responded more quickly than control participants, although this contrast was only marginally reliable - In the 2-back condition, there was no difference between these conditions
Ancillary Measures: - As in Experiments 1 through 3, participants rated the antisaccade task as more difficult than the prosaccade task - As expected, the 2-back condition yielded higher task difficulty ratings than did the 0-back condition - Participants in the 2-back condition reported being more concerned with the experimenter's evaluation than did participants in the 0-back condition - Also, participants in the 2-back condition reported lower levels of effort than did participants in the 0-back condition - Finally, participants in the 2-back condition did not expect to perform as well as participants in the 0-back condition - No other effects were reliable.

Was the precision of the estimate of the intervention or treatment effect reported?

- CONSIDER:

– Were confidence intervals (CIs) reported?

☐ Yes
☒ No
☐ Can't tell

Are there any obvious shortcomings in the reporting of the data?

☐ Yes (please specify)
☒ No

Do the authors report on all variables they aimed to study as specified in their aims/research questions?

This excludes variables just used to describe the sample.

☒ Yes (please specify)
☐ No

Do the authors state where the full original data are stored?

☐ Yes (please specify)
☒ No

What do the author(s) conclude about the findings of the study?

Please give details and refer to page numbers in the report of the study where necessary.

☐ Details

Experiment 1: Consistent with the success of the stereotype threat manipulation, participants in the stereotype threat condition reported that there were gender differences in the performance of the task and that male performance was superior to that of female performance. Although male performance was unaffected by this manipulation, women in the stereotype threat condition were less accurate in their identification of target orientation than women in the control condition and men in either condition.
This finding is consistent with the working memory inference account of stereotype threat effects. However, this finding is not necessarily inconsistent with the mere effort account.

Experiment 2: Once again, the manipulation checks suggest that the stereotype threat manipulation was successful. Participants exposed to the stereotype threat manipulation believed that the saccade tasks were diagnostic of their math ability and that these tasks were gender biased.
In Experiment 2, participants in the stereotype threat condition were as accurate in their identifications as participants in the control group, and reported target orientation reliably more quickly. This finding is inconsistent with an explanation that relies solely on working memory inference, which would predict that participants subject to threat should perform more poorly than controls on the antisaccade task at all target display times.

Experiment 3: Women subject to stereotype threat responded to target orientation as accurately but more quickly than controls. The eye tracking data allowed us to examine

the processes that produced this outcome. Consistent with the predictions of both the working memory interference and mere effort accounts, on the antisaccade task, participants subject to the threat manipulation launched more reflexive saccades than controls. On these trials, the threat group also launched corrective saccades more quickly than controls. In addition, on trials on which participants were successful in inhibiting the tendency to launch a reflexive saccade, participants subject to stereotype threat launched correct saccades faster than participants in the control group. These findings are consistent with the mere effort account but are inconsistent with the working memory interference account. Finally, finding faster adjusted reaction times for stereotype threat than control participants is not predicted by the working memory interference account but is consistent with the mere effort account.

Experiment 4:

Replicating the findings of Experiments 2 and 3, in the 0-back condition, stereotype threat participants responded more quickly than control participants, as would be expected when working memory resources are available. However, when working memory resources were consumed by the 2-back task, threat participants performed more poorly than control participants. Thus, adding a high-load concurrent task that taxed working memory debilitated the performance of participants subject to threat on the antisaccade task, whereas stereotype threat alone facilitated performance when there was no (see Experiments 2 and 3) or a minimal cognitive load (0-back, Experiment 4).

It is also interesting to note that the high-load concurrent task manipulation did not impair the performance of control participants. In fact, participants in the control group responded more quickly in the 2-back than in the 0-back condition and were equally accurate.

General Discussion:

Consistent with the notion that stereotype threat impairs working memory capacity, Experiment 1 demonstrated that when the display time was 150 ms, participants subject to threat performed more poorly than controls. However, when display time was increased to 250 ms in Experiment 2, stereotype threat participants performed better than controls. This finding is inconsistent with an explanation that relies on working memory interference alone, which suggests that participants subject to stereotype threat should perform more poorly than control participants regardless of display time.

However, the findings of Experiments 1 and 2 are consistent with the account provided by mere effort (Harkins, 2006).

In Experiment 3, we examined the participants' eye movements during task performance to identify the processes that resulted in the terminal performances outcomes of Experiments 1 and 2. Consistent with the mere effort account, participants subject to stereotype threat launched more reflexive saccades, but they also generated corrective and correct saccades more quickly than participants in the control condition. Also consistent with the mere effort account was the finding that the adjusted reaction times for participants subject to threat were faster than those for control participants.

The results of Experiments 1–3 are consistent with the operation of the single process suggested by the mere effort account. In Experiment 4, we asked participants to complete the antisaccade task while also performing a 2-back concurrent task. When stereotype threat participants completed the antisaccade task under low-working memory load (0-back), they performed better than control participants, just as they did in Experiments 2 and 3, but

under high-working memory load (2-back), they performed more poorly. These findings show that working memory resources are required for threat participants to outperform the control participants in Experiments 2 and 3.

On the basis of previous research on the effects of reduced working memory capacity on performance on the antisaccade task (Kane et al., 2001; Roberts et al., 1994; Stuyven et al., 2000; Unsworth et al., 2004), the working memory interference account predicted that threat should degrade performance on this task regardless of target display time. This prediction was not supported, whereas the findings are consistent with the mere effort account.

Research conducted subsequent to Schmader and Johns's (2003) work that has invoked the concept of working memory interference is also consistent with the mere effort account.

Although the present findings are consistent with the mere effort account, there are a number of factors that limit the range of conditions under which we would expect these effects to be produced.

Thus, our research shows how threat participants perform when they find themselves in a situation that they did not seek out and that does not provide a plausible explanation as to why they may not perform well. Under these circumstances, they are motivated to perform well. As a result, the mere effort account is not inconsistent with research that suggests that instead of seeking out the opportunity to demonstrate how well they can perform, participants under stereotype threat are quite willing to take advantage of explanations that will allow them to deflect responsibility for their performance (e.g., Ben-Zeev et al., 2005; Steele & Aronson, 1995; Stone, 2002).

Quality of the study - Reporting

Is the context of the study adequately described?

Consider your answer to questions: Why was this study done at this point in time, in those contexts and with those people or institutions? (Section B question 2) Was the study informed by or linked to an existing body of empirical and/or theoretical research? (Section B question 3) Which of the following groups were consulted in working out the aims to be addressed in the study? (Section B question 4) Do the authors report how the study was funded? (Section B question 5) When was the study carried out? (Section B question 6)

☐ Yes (please specify)
☐ No (please specify)

Are the aims of the study clearly reported?

Consider your answer to questions: What are the broad aims of the study? (Section B question 1) What are the study research questions and/or hypotheses? (Section C question 10)

☐ Yes (please specify)
☐ No (please specify)

Is there an adequate description of the sample used in the study and how the sample was identified and recruited?

Consider your answer to all questions in Methods on 'Sampling Strategy', 'Recruitment and Consent', and 'Actual Sample'.

☐ Yes (please specify)
☐ No (please specify)

Is there an adequate description of the methods used in the study to collect data?

Consider your answer to the following questions in Section I: Which methods were used to collect the data? Details of data collection methods or tools Who collected the data? Do the authors describe the setting where the data were collected? Are there other important features of the data collection procedures?

☐ Yes (please specify)
☐ No (please specify)

Is there an adequate description of the methods of data analysis?

Consider your answer to the following questions in Section J: Which methods were used to analyse the data? What statistical methods, if any, were used in the analysis? Who carried out the data analysis?

☐ Yes (please specify)
☐ No (please specify)

Is the study replicable from this report?

☐ Yes (please specify)
☐ No (please specify)

Do the authors avoid selective reporting bias?

(e.g. do they report on all variables they aimed to study as specified in their aims/research questions?)

☐ Yes (please specify)
☐ No (please specify)

Quality of the study - Methods and data

Are there ethical concerns about the way the study was done?

Consider consent, funding, privacy, etc.

☐ Yes, some concerns (please specify)
☐ No concerns

Were students and/or parents appropriately involved in the design or conduct of the study?

☐ Yes, a lot (please specify)
☐ Yes, a little (please specify)
☐ No (please specify)

Is there sufficient justification for why the study was done the way it was?

☐ Yes (please specify)
☐ No (please specify)

Was the choice of research design appropriate for addressing the research question(s) posed?

☐ Yes (please specify)
☐ No (please specify)

To what extent are the research design and methods employed able to rule out any other sources of error/bias which would lead to alternative explanations for the findings of the study?

e.g. (1) In an evaluation, was the process by which participants were allocated to or otherwise received the factor being evaluated concealed and not predictable in advance? If not, were sufficient substitute procedures employed with adequate rigour to rule out any alternative explanations of the findings which arise as a result? e.g. (2) Was the attrition rate low and if applicable similar between different groups?

☐ A lot (please specify)
☐ A little (please specify)
☒ Not at all (please specify)

How generalisable are the study results?

☐ Details

- not at all

Weight of evidence - A: Taking account of all quality assessment issues, can the study findings be trusted in answering the study question(s)?

In some studies it is difficult to distinguish between the findings of the study and the conclusions. In those cases please code the trustworthiness of this combined results/conclusion. ***Please remember to complete the weight of evidence questions B-D which are in your review specific data extraction guidelines.***

☐ High trustworthiness (please specify)
☒ Medium trustworthiness (please specify)
☐ Low trustworthiness (please specify)

Have sufficient attempts been made to justify the conclusions drawn from the findings so that the conclusions are trustworthy?

☐ Not applicable (results and conclusions inseparable)

☐ High trustworthiness

☐ Medium trustworthiness

☒ Low trustworthiness

- To me it seems like the authors want to jump to the conclusion that the mere effort account explains the results. For example "These findings suggest that participants saw the target at the same time, at the 400-ms mark. Thus, it was the motivation to press the key to make the response that produced the reaction time difference between the conditions on the prosaccade trials", for me it does not seem like this is the only explanation for the results but the authors jump there instantaneously.
- Furthermore I am not sure whether or not the authors theory (mere effort account) is actually different than the working memory inference account. It might just not mention which parts of the working memory are actually being used and to instead claim that the mere effort account explains everything seems a bit premature and too easy.

Wells et al. (2014)

CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

Is the case definition adequate?

- a) yes, with independent validation
- b) yes, e.g., record linkage or based on self reports
- c) no description

Representativeness of the cases

- a) consecutive or obviously representative series of cases *
- b) potential for selection biases or not stated

Selection of Controls

- a) community controls *
- b) hospital controls
- c) no description

Definition of Controls

- a) no history of disease (endpoint) *
- b) no description of source

Comparability

Comparability of cases and controls on the basis of the design or analysis

- a) study controls for _____ (Select the most important factor.) *
- b) study controls for any additional factor * (This criterion could be modified to indicate specific control for a second important factor.)

Exposure

Ascertainment of exposure

- a) secure record (e.g., surgical records) *
- b) structured interview where blind to case/control status *
- c) interview not blinded to case/control status
- d) written self report or medical record only
- e) no description

Same method of ascertainment for cases and controls

- a) yes *
- b) no

Non-Response rate

- a) same rate for both groups *
- b) non respondents described
- c) rate different and no designation

COHORT STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability.

Selection

Representativeness of the exposed cohort

- a) truly representative of the average _____ (describe) in the community *
- b) somewhat representative of the average _____ in the community *

- c) selected group of users, e.g., nurses, volunteers
- d) no description of the derivation of the cohort

Selection of the non exposed cohort

- a) drawn from the same community as the exposed cohort *
- b) drawn from a different source
- c) no description of the derivation of the non exposed cohort

Ascertainment of exposure

- a) secure record (e.g., surgical records) *
- b) structured interview *
- c) written self report
- d) no description

Demonstration that outcome of interest was not present at start of study

- a) yes *
- b) no

Comparability

Comparability of cohorts on the basis of the design or analysis

- a) study controls for _____ (select the most important factor) *
- b) study controls for any additional factor * (This criterion could be modified to indicate specific control for a second important factor.)

Outcome

Assessment of outcome

- a) independent blind assessment *
- b) record linkage *
- c) self report
- d) no description

Was follow-up long enough for outcomes to occur

- a) yes (select an adequate follow up period for outcome of interest) *
- b) no

Adequacy of follow up of cohorts

- a) complete follow up - all subjects accounted for *
- b) subjects lost to follow up unlikely to introduce bias - small number lost - > _____ % (select an adequate %) follow up, or description provided of those lost) *
- c) follow up rate < _____% (select an adequate %) and no description of those lost
- d) no statement

University of Glasgow (n.d.)

DOES THIS REVIEW ADDRESS A CLEAR QUESTION?

Did the review address a clearly focussed issue?

- Was there enough information on:
  - The population studied
  - The intervention given
  - The outcomes considered

☐ Yes
☐ Can't tell
☐ No

Did the authors look for the appropriate sort of papers?

- The 'best sort of studies' would:
  - Address the review's question
  - Have an appropriate study design

☐ Yes
☐ Can't tell
☐ No

ARE THE RESULTS OF THIS REVIEW VALID?

Do you think the important, relevant studies were included?

- Look for:
  - Which bibliographic databases were used
  - Follow up from reference lists
  - Personal contact with experts
  - Search for unpublished as well as published studies
  - Search for non-English language studies

☐ Yes
☐ Can't tell
☐ No

Did the review's authors do enough to assess the quality of the included studies?

- The authors need to consider the rigour of the studies they have identified. Lack of rigour may affect the studies results.

☐ Yes
☐ Can't tell
☐ No

If the results of the review have been combined, was it reasonable to do so?

- Consider whether:
  - The results were similar from study to study
  - The results of all the included studies are clearly displayed

  - The results of the different studies are similar
  - The reasons for any variations are discussed

☐ Yes
☐ Can't tell
☐ No

WHAT ARE THE RESULTS?

What is the overall result of the review?

- Consider:
  - If you are clear about the review's 'bottom line' results
  - What these are (numerically if appropriate)
  - How were the results expressed (NNT, odds ratio, etc)

How precise are the results?

- Are the results presented with confidence intervals?

☐ Yes
☐ Can't tell
☐ No

WILL THE RESULTS HELP LOCALLY?

Can the results be applied to the local population?

- Consider whether:
  - The patients covered by the review could be sufficiently different from your population to cause concern
  - Your local setting is likely to differ much from that of the review

☐ Yes
☐ Can't tell
☐ No

Were all important outcomes considered?

☐ Yes
☐ Can't tell
☐ No

Are the benefits worth the harms and costs?

- Even if this is not addressed by the review, what do you think?

☐ Yes
☐ Can't tell
☐ No

References

Critical Appraisal Skills Programme. (2018). CASP Systematic Review Checklist [Organization]. In *CASP - Critical Appraisal Skills Programme*. https://casp-uk.net/casp-tools-checklists/.

EPPI-Centre. (2003). *Review guidelines for extracting data and quality assessing primary studies in educational research* (Guidelines Version 0.9.7). Social Science Research Unit.

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, *93*(4), 544–564. https://doi.org/10.1037/0022-3514.93.4.544

University of Glasgow. (n.d.). *Critical appraisal checklist for a systematic review* [Checklist]. Department of General Practice, University of Glasgow.

Wells, G., Shea, B., O'Connell, D., Robertson, J., Welch, V., Losos, M., & Tugwell, P. (2014). The newcastle-ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. *Ottawa Health Research Institute Web Site*, *7*.