Non-linear effects of stress on eyewitness memory

AUTHORS:
Milton Gering[1]
Tayla Johnson[1]
Colin Tredoux[1]

AFFILIATION:
[1]Department of Psychology, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Milton Gering

EMAIL:
miltongering@gmail.com

DATES:
Received: 27 Aug. 2021
Revised: 12 Sep. 2022
Accepted: 18 Oct. 2022
Published: 29 Mar. 2023

HOW TO CITE:
Gering M, Johnson T, Tredoux C. Non-linear effects of stress on eyewitness memory. S Afr J Sci. 2023;119(3/4), Art. #12102. https://doi.org/10.17159/sajs.2023/12102

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from author(s)
☐ Not available
☐ Not applicable

EDITOR:
Leslie Swartz

KEYWORDS:
stress, eyewitness, memory, line-up, faces

FUNDING:
None

The effect of stress on face identification is not yet clear, with recent experiments finding positive, negative and null results. Here we report the results of two experiments examining the effect of stress on eyewitness performance in line-up face recognition tasks. Both experiments use a stress manipulation and live mock crime paradigm to examine the relationship between stress at encoding and subsequent line-up performance. Experiment 1 replicated an experiment by Sauerland et al. (Behav Sci Law. 2016;34(4):580–594) which induced stress using the Maastricht Acute Stress Test. The replication found the same null result as the original experiment. Experiment 2 aimed to address a limitation of many laboratory experiments which dichotomise stress into low and high groups for comparison. As the Yerkes-Dodson law (1908) suggests that a non-linear relationship exists between stress and performance, it was hypothesised that using a low, medium and high stress manipulation might show clearer results than a dichotomous paradigm. The results of Experiment 2 show a non-linear relationship, with no difference between the low and high stress groups but better performance by the middle stress group. The results suggest that a different approach is required in experiments on stress and face recognition, as the stress–performance relationship is likely non-linear.

Significance:

- Non-linear models are better predictors of face recognition in line-up tasks than are linear models.
- Two group designs provide insufficient resolution to capture the stress–performance relationship.

Non-linear effects of stress on eyewitness memory

People who witness crimes often experience high levels of stress during the event. The effect that this stress might have on memory is important, yet remains unclear. Although an earlier meta-analysis[1] showed a negative effect of stress at encoding, recent studies have found mixed results, and a more recent meta-analysis concludes that extant evidence does not show that stress affects witness memory[2]. Some of the differences may be a result of different methods, particularly for stress induction, used in studies within the literature, but it is likely also that the stress induction paradigms used in the corpus of studies have produced unreliable results. Where the Deffenbacher review found a negative linear effect[1], the more recent review suggests that a non-linear model may offer a better explanation[2]. We report on two studies that show that a dichotomous stress manipulation will not achieve reliable results, especially because the stress–performance relation is explained by the universally cited Yerkes-Dodson law[3] to be non-linear (likely quadratic). To see a non-linear effect, it is necessary to induce a minimum of three levels of stress. In doing so, the studies presented here show a flaw in the existing literature and describe an experimental solution.

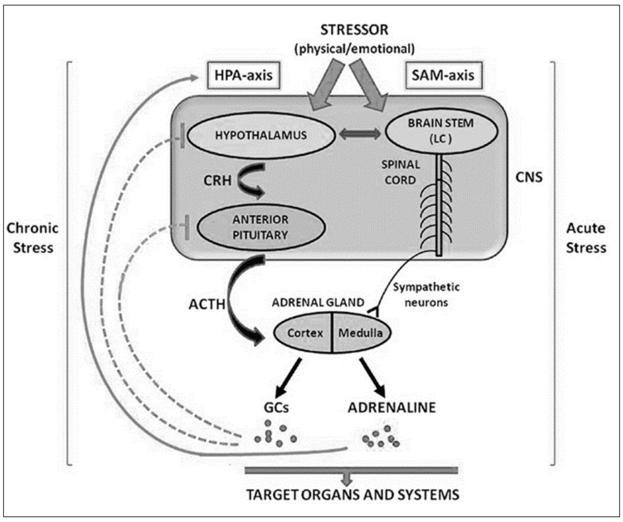
Stress and memory

Acute episodes of stress have known physiological consequences, primarily preparing the body for action[4], but simultaneously having effects on memory. These responses occur through the activation of two systems: the fast-acting sympathetic-adrenal-medullary (SAM) axis, or the slower-acting hypothalamic-pituitary-adrenal (HPA) axis (see Figure 1). The SAM axis response increases activity of the sympathetic nervous system, releasing adrenaline, which in turn stimulates the release of norepinephrine in the brain.[5] The norepinephrine activates the amygdala, which interacts with the brain regions involved in encoding memories, including the hippocampus and frontal lobes.[4] These areas are also activated through the HPA axis, their glucocorticoid receptors responding to the release of a different hormone, cortisol.[5] This complex response is experienced by many witnesses of crimes and may affect how they process and encode events. The combined cortisol and noradrenaline response is thought to affect memory, impacting different brain regions involved in perceiving, encoding and storing memories for later recall and retrieval.

Several factors influence the effect of acute stress on memory, including the amount of stress experienced, the degrees of activation of the SAM and HPA axes, the types of memory process, and the time since onset of stress.[6] While both axes activate the amygdala and hypothalamus, SAM activation appears to boost memory consolidation, while HPA axis activation weakens it.[7] These counteracting paths may explain why mild or moderate stress during encoding improves memory, yet intense stress impairs it.[8] This non-linear effect of stress on memory has been known experimentally since 1908[3] and is often referred to as the 'Yerkes-Dodson law', shown here as Figure 2. The quadratic curve shown in the figure may not be as symmetrical as suggested, perhaps having a steep drop off at extreme levels of stress ('catastrophic forgetting').[9] Despite some scholars contesting the Yerkes-Dodson law, or its general application, there has been no empirical work with the type of stress inductions needed to falsify the theory.[2]

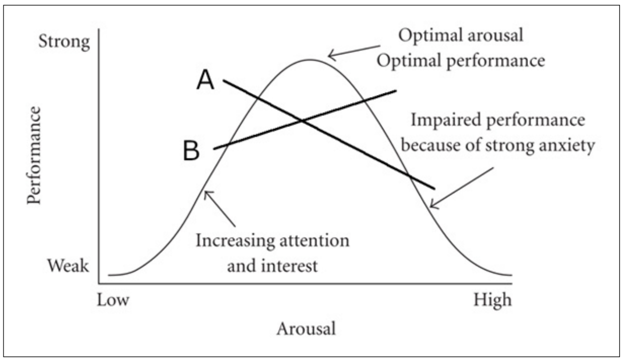
A significant limitation in the extant literature on stress and witness memory is that almost all studies use a two-group design (typically 'high' and 'low' stress groups), despite the obvious impossibility of detecting a non-linear effect with two levels of a predictor. This may be why different stress responses are seen, as a dichotomous division of a non-linear variable will show different patterns depending on the intensity of stress induced.

In understanding the effects of stress on witness memory, it is also important to note that, to remember details of an event, one must notice and attend to it. Attentional resources are limited, and not everything in the environment can be attended to. Emotional events, such as fear-inducing crimes, tend to be conspicuous and thus attract attention, particularly to central, life-threatening aspects of the event, leading to poorer encoding of other information.[3] Arousal will also impair memory in situations where cognitive demands are high, as the brain must process the stress response while simultaneously using other cognitive resources.[10] This may affect how successfully different information is encoded by witnesses into their memory.



Source: Diamond et al.[21] licensed under a Creative Commons licence

Figure 1: Schematic representation of the actions of the hypothalamic-pituitary-adrenal (HPA) and sympathetic-adrenal-medullary (SAM) axes in the regulation of the stress response.



Source: Baritaki et al.[22] licensed under a Creative Commons licence

Figure 2: The Yerkes-Dodson Curve. The lines A and B show how it is possible to find negative and positive (but misleading) effects of stress on performance. The Yerkes-Dodson curve could be simply quadratic, or taper off at the end to look more Gaussian, or be asymmetrical and piecewise and only roughly approximate a quadratic curve.

Another factor to consider is the nature of the memory tasks required of eyewitnesses to crimes. Witnesses to a crime are required to *recall* events, and to *recognise* perpetrators, which are explicit and implicit memory tasks, respectively.[11] Laboratory studies of memory show that stress at encoding benefits recall tasks more than recognition tasks.[12] Similar task-dependent differences have been seen in studies on depression. While depression is generally considered an affective disorder, it also produces changes in cognition.[13] Participants with major depressive disorder perform worse than healthy controls in explicit memory tasks.[14] This suggests that depression may impair a witness's ability to recall events, and probably more than it would impair recognition of perpetrators. As such, it should be measured and included as a control variable. While mood disorders may be a source of confounding influences in the literature, whether the memory task requires implicit or explicit memory appears to be an overarching moderator of the stress–performance relationship.

We report two studies here that investigate the effect of stress on witness memory, simulating a criminal event as part of a live interaction. We aimed to capture the essence of a stressful crime by inducing stress and presenting a critical event on which participants were later tested. Both these studies include elements considered best practice in the stress induction literature, including a meaningful delay between encoding and recognition, the use of physiological and self-report manipulation checks of stress induction, and well-constructed line-ups for testing recognition memory.[15,16] We believe that these methodological elements are important parts of the three-group design that is critical for mapping the non-linear relationship between stress and performance. Although we consider both recall and recognition memory, our analysis is focused mostly on participant recognition of perpetrators from line-ups.

Method

Design

Both studies used an experimental design to investigate the effect of induced stress on eyewitness memory. In Experiment 1, stress was manipulated, following common practice in the literature[15], into low and high levels of stress, and effects tested on perpetrator recognition and event recall. For Experiment 2, stress was manipulated to have three levels: a control group had no intentional stress induction, one experimental group received a moderate stress induction, and a second experimental group received a high stress induction.

Participants

For both experiments, participants were recruited via the Student Research Participation Programme (SRPP) run by the Department of Psychology at the University of Cape Town (UCT). Participants were recruited via an advertisement on a university online portal. For Experiment 1, 180 students were screened for eligibility and 123 of those were invited to take part in the laboratory session. Of those 123 participants, 40 (8 men) took part in the 45-min laboratory session, completing the Maastricht Acute Stress Test (MAST), with 28 participants (6 men) providing a full set of data. As such, there were 40 participants in the stress induction phase and 28 in the eyewitness simulation phase. For Experiment 2, 190 (of 266) participants were invited to take part in the laboratory phase of the experiment after screening. Of those, 89 took part in the experiment. Of these, 14 were previously familiar with the 'perpetrator' and so were excluded from analysis, leaving 75 participants.

Screening instruments

Various instruments were used to screen, check manipulations, or measure covariates. These processes were undertaken through an online survey in the weeks prior to the in-person laboratory sessions, and, as they are not central to our hypotheses, we mention them only briefly, as follows (values in parentheses are Cronbach alpha estimates of internal consistency, or test-rest reliability): The 4-item Primary Care Post-Traumatic Stress Disorder Screen (PC-PTSD; $\alpha=0.89$; $r=0.83$), the State-Trait Anxiety Inventory Form Y1 and Y2 (STAI-Trait and STAI-State; $\alpha=0.89$; $r=0.69$–0.89), Beck Depression Inventory-II (BDI-II; $\alpha=0.90$; $r=0.93$), the Beck Anxiety Inventory (BAI; $\alpha=0.92$; $r=0.75$), the Positive and Negative Affect Schedule (PANAS; $\alpha=0.89$; $r=0.83$).

In both experiments we excluded students who had been diagnosed with a psychiatric disorder, who had high levels of trait anxiety or a history of head injury, and who had been affected by psychological trauma. Anyone answering 'yes' to the head trauma question or any of the four trauma questions, or scoring in the 'severe' category for anxiety or depression was not invited to participate in the in-person laboratory session. An additional exclusion criterion was used for Experiment 2, namely students who had a history of plagiarism at UCT were excluded. Only young adults, between 18 and 25, were invited to participate in the studies.

Experimental materials

The Maastricht Acute Stress Test

The MAST is a stress induction procedure that combines the Cold Pressor Test with mental arithmetic from the Trier Social Stress Test.[17] Participants alternate between immersing a hand in cold water and doing an arithmetic task in which they must count backwards in steps of 17 from 2043. The task is known to activate both the HPA and SAM axes, eliciting a comprehensive stress response.[17] This procedure was used to induce stress in Experiment 1, having previously been used in a similar experiment in Maastricht.[16] The control group goes through the Placebo MAST, which uses warm water and a simple counting task so as not to induce stress but to maintain the equivalent duration of the procedure.

The Self-Administered Interview

The Self-Administered Interview (SAI©) is a standardised, self-administered interview, based on the cognitive interview.[18] The interviewer asks generic questions applicable to various crimes; this test has been shown to enhance recall[18] and was used in Experiment 1.

The Vrije Universiteit Ambulatory Monitoring System

Heart rate (HR) and skin conductance (SC) were measured using the Vrije Universiteit Ambulatory Monitoring System (VU-AMS) in both experiments as part of the stress manipulation. Both HR and SC have been shown to increase when people experience stress.[18]

Line-ups

Target absent (TA) and target present (TP) colour line-ups were constructed for each target, all of whom were volunteers studying at UCT. Both experiments used two unique targets and so two separate TA and TP line-ups were constructed for each experiment. All line-ups consisted of six different faces photographed from the neck upwards against a pale grey background. Each individual in the line-up wore a black t-shirt. For Experiment 1, foils were selected to be similar to the target; for Experiment 2, foils were selected to match a verbal description of the target. These methods are both recommended in the literature and require similarity to be judged by the line-up constructor. Line-ups were constructed by the first author (M.G.) for Experiment 1 and the second author (T.J.) for Experiment 2, before being assessed for similarity and 'pop-out' effects by the study supervisor. Two versions of each line-up were constructed, with targets in different positions, to reduce the influence of a position effect. All targets and foils were women in their 20s; three were Caucasian and one (in Experiment 1) was East Asian. Photographs of foils were retrieved from the supervising author's database of line-up materials.

Procedure

Ethical clearance for both experiments was obtained through the Department of Psychology Ethics Committee at UCT.

Screening phase, Experiment 1

The STAI-Trait and PC-PTSD inventories were used to screen out (and thus protect) participants with high anxiety and those with symptoms of PTSD. These tests were administered online.

Laboratory phase, Experiment 1

Once consent forms had been signed, participants attached the VU-AMS device to themselves, and it was set to record. Participants then completed the BDI-II which was included in the modelling as a control. Participants were informed of the procedure for the MAST, or placebo-MAST for the controls, before commencing with the test. For the experimental group, the water was cooled with ice to 5 °C and for the control group it was warmed to 20 °C. After the MAST procedure, participants' state anxiety was measured with the STAI-State. Thereafter, participants were given a vocabulary test as a distractor task but told that it was the main test. This was done so as not to alert participants that the main test was in fact the mock crime. At this point, the experimenter left the room, leaving a phone on the VU-AMS container. A confederate then knocked on the door, entered the room and told the participant that they had left their phone behind during a previous session. The confederate then went to the table, took the researcher's phone, and left the room. Confederates always wore blue jeans and simple tops in dull colours.

When the researcher reappeared, he asked about the distractor task while searching for the phone, noted that the phone was missing and asked participants if they had seen anything. Once participants had told the researcher what had happened, they were informed that they had witnessed a mock crime and were debriefed. An email containing both a recall and recognition task, ensuring blind administration, was sent to participants the next day. A response period of 24 hours ensured that participants were only affected by stress during encoding, an important control (cf. Sauerland et al.[16]). The task sent to participants consisted of the SAI as well as two line-ups. Each participant first saw a TA line-up, and then a TP line-up, for the confederate/thief they saw in the lab. The repeat measure was used to increase power, given the small sample size. Both line-ups had unique foils, and participants did not know to expect two line-ups or that one would not include the target, thus mimicking real-world uncertainty. Instructions stated that 'the suspect may or may not be present in the line-up'. Participants were asked to rate their confidence in their line-up selection and had the option to say the suspect was not present or that they did not know who the suspect was. The information from the SAI was used to code for accuracy of descriptions by the witnesses. Confidence was not analysed in the present study.

Screening phase, Experiment 2

The same screening procedure was used as in Experiment 1, although the STAI was replaced with the BAI.

Laboratory phase, Experiment 2

Participants arrived for in-person testing group sessions with a maximum of 10 participants, only one of whom was asked to wear the VU-AMS. Each group was assigned to one of three stress conditions (control, moderate stress, high stress). Upon arrival at the laboratory, participants were seated in front of a computer with a consent form and a demographic information sheet, and reported their negative affect state using the PANAS. Once forms were completed, the participants selected for physiological monitoring attached themselves to the VU-AMS device. Participants were told that while they completed a task, a research assistant would be checking their academic plagiarism history. Each stress group received different instructions on this task: the control group directly copied the extract, the moderate stress group summarised it, and the high stress group summarised the extract but were told that it would be marked and compared with the rest of the group's. After this task, the first target emerged from a curtained area in the room and delivered a false plagiarism report: the control group was told their submissions were free of plagiarism; the moderate stress group was sternly told that plagiarism was present in some of the submissions but no further action was to be taken; and the high stress group was sternly told that plagiarism was present, that participants would be identified on their way out of the session, and that they were to appear in a student disciplinary tribunal to plead their case at a later date. This was the key point for the stress induction. Target one then went back into the curtained area after 90 seconds of exposure and interaction with participants.

Participants again reported their affective state, and the researcher informed participants that the plagiarism report was false, before giving them a distractor task. The second target was then introduced by the researcher. The target emerged from a curtained area in the room for 90 seconds and debriefed the participants on what the task had really consisted of. The targets were assigned to roles in counter-balanced order, across groups, which, along with the randomisation of experimental condition, was done through MS Excel. The researcher collected a final measure of affective state before debriefing participants. Participants were told that they would be required to identify the two targets in a photographic line-up the following day and were emailed a copy of the debriefing form and a link to two photographic line-ups 24 hours after their laboratory participation. Participants were instructed to identify the two targets to whom they were exposed previously and given the option to indicate if they did not know if the target was present, or to indicate that the target was not present in the line-up. The identifications were done online using simultaneous photographic line-ups and contained a TA line-up for one of the targets, and a TP line-up for the other target. Participants always saw one TA and one TP line-up.

Data management and statistical analysis

For both studies, data were analysed with the R statistical programming language[19], and the lme4 package[20]. All assumptions for analysis were met unless otherwise stated.

In Experiment 1, HR and SC were sampled between minutes two and three to obtain a baseline measurement of these physiological measures of stress. A second measure was taken between minutes 16 and 17 of the recording, by which time the MAST had been completed. A mixed-methods ANOVA was conducted to see if there were significant main or interaction effects between experimental manipulations and time. The STAI-State was administered in the second time interval, post-MAST, and a t-test was conducted to see if there was a significant difference in STAI-State between groups post-stress induction. A list of verifiable details was coded for details of the mock crime, descriptions of the perpetrator and the location of the event. Inter-rater reliability was obtained for the 28 coded statements on details of location, events and perpetrator descriptions. Descriptions were generated by the first author and checked with the confederates as in the original study we replicate[16] before being checked by a research assistant who was the second coder. Intra-class inter-rater reliability coefficients were 0.89, 0.93 and 0.82 for location, events and perpetrator details, respectively. An analysis of covariance (ANCOVA) was conducted for each category of recall, and on total details recalled, to analyse the effect of condition (stress vs control) on recall memory, using BDI-II as a covariate. For face identification, a logistic regression was conducted to compare the results between the stress and control groups for both TP and TA line-ups.

For Experiment 2, the negative affect score was calculated by using the negative items from the PANAS at the three time points at which physiological measures were also taken. A 3x3 mixed-designs ANOVA was conducted to ascertain whether a significant interaction occurred between the three stress groups at the different time intervals in self-report measures of negative affect. Furthermore, a planned contrast to check for the predicted differences between groups was run on negative affect measurements taken directly after the stressor to see whether the data matched the expected pattern of low, moderate and high reported stress. Once the raw data for line-up responses had been tabulated, the variables for each target line-up were reduced and coded as dichotomous variables with levels accurate and inaccurate (i.e. correct rejections in TA line-ups and correct identifications in TP line-ups were coded as accurate). It is important to note that 'don't know' responses were coded as 'inaccurate' for line-ups. A separate quadratic logistic regression was then run for each target in R to ascertain whether stress predicted accuracy in suspect identification in the expected inverted-U shape.[8]

Results

Experiment 1

Manipulation check

HR and SC were analysed for 35 participants (experimental group = 17) using a repeated measures ANOVA to check the stress manipulation on physiological responses. The STAI-Y1 scores of 40 participants (experimental group = 20) were used as a self-report manipulation check on psychologically experienced stress. Both groups completed the STAI-Y1 after doing the MAST or placebo MAST (for the control group). An independent samples t-test was conducted to see if there was a significant difference between the control and high stress groups post-MAST. Results are summarised in Table 1. Figure 3 shows the interaction effects of experimental group and time of measurement for each of HR and SC. This interaction between group and time on physiological measures shows opposite patterns in arousal for the stress and control groups. These results indicate that the stress induction procedure was successful.

Effect of stress on recall

In Experiment 1, recall data were collected from 28 participants (stress group = 14). Means and standard deviations are reported in Table 2 for each category of recall (events, location and participant description), and a total score for each group is also reported. Comparing the results for total recall across conditions suggests that the MAST group performed slightly worse than the control group. Separate ANCOVAs were conducted for memory of location, events and descriptions using depression as a covariate. For the analysis for memory of events, the more robust Welch was conducted to handle the homogeneity of variance assumption. The *p*-values and effect sizes for the ANCOVAs are reported alongside the descriptive statistics.

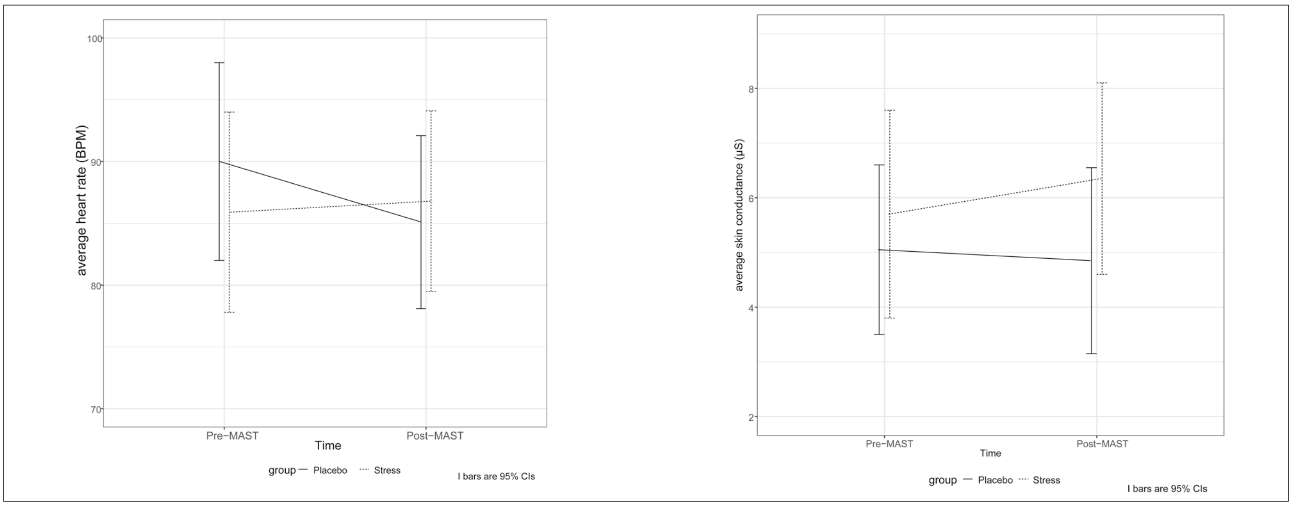


Figure 3: Disordinal heart rate x time and skin conductance x time interactions for the low (Placebo MAST) and high (MAST) stress groups in Experiment 1, showing averages and confidence intervals.

Table 1: Self-reported stress and depression scores by condition

	STAI-Y1			BDI(II)			Heart rate			Skin conductance		
	M	SD	d	M	SD	d	M	SD	d	M	SD	d
MAST	24.35	9.97	0.95	13.90	9.70	0.42	0.90	10.47	0.84*	0.20	2.25	0.92*
Placebo	14.30	8.94		10.35	6.45		-0.42	10.37		0.65	2.45	

Note: Means (M), standard deviations (SD) and effect sizes for manipulation checks. For heart rate and skin conductance, mean difference is reported.

**Indicates* $p<0.05$ *for the inferential test from which the effect size was produced.*

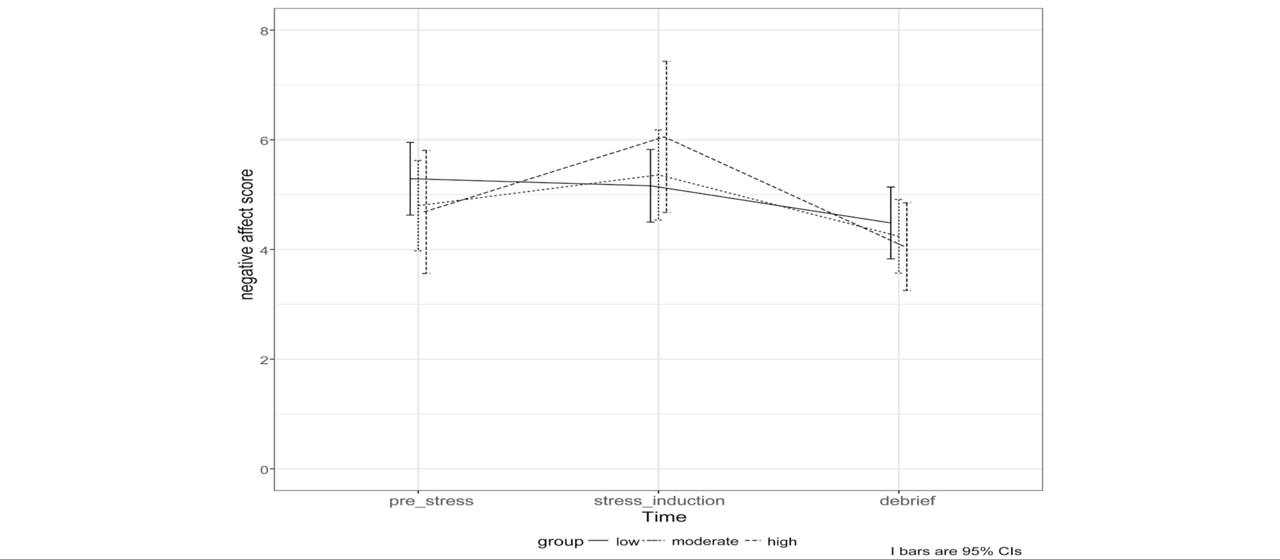


Figure 4: Time x group interaction of total negative affect scores from the PANAS, Experiment 2.

Table 2: Means and standard deviations for recall type by condition

	MAST		Placebo		ANCOVAs		
	M	SD	M	SD	F	*p*	d
Location details	4.86	3.08	6.50	2.98	1.85	0.190	0.54
Event details	5.21	2.63	5.21	1.19	0.01	0.934	0.03
Perpetrator description	3.93	2.37	4.36	2.92	0.12	0.646	0.18
Total	14.00	5.94	16.07	3.87	0.71	0.441	0.29

Note: Means (M) and standard deviations (SD) for each recall type for the stress and control group.

Effect of stress on face identification

Face identification data were collected from 28 participants (stress group = 14). From Table 3, which shows the response data for the face identifications, we can see that there is no difference between the groups. Of the 28 participants, 15 made identification attempts in the TP line-up across the groups, and 8 participants were successful. The results of the logistic regression show no difference between the high and low stress groups on TP (B=0.04, SE=1.10, $p=0.974$) or TA (B=0.97, SE=1.34, $p=0.480$) line-ups with no effect of depression on either task ($p=0.718$ and $p=0.492$, respectively). Thus, we can conclude that there were no meaningful differences in identification rate between the groups.

Table 3: Line-up responses by stress condition for target-absent and target-present line-ups

	MAST	Placebo
Target absent	*n*=14	*n*=14
Correct reject	3	2
False alarm	10	9
'Don't know' responses	1	3
Target present	*n*=14	*n*=14
Correct identification	4	4
Foil identification	3	4
False reject	5	6
'Don't know' responses	2	0

Note: For a target-absent line-up, rejecting the line-up is the best response, equivalent to a correct identification in the target-present line-up. False alarm in a target-absent line-up and foil identification are also equivalent.

Experiment 2

Manipulation check

HR was obtained and analysed from 10 participants (one from each session): three in the control group; three in the medium-stress group; and four in the high-stress group. The mean change values (see Table 4) indicate that HR increased for all groups throughout the experiment. However, due to the small sample for whom we collected physiological measures due to limitations of time and equipment, and the ensuing low statistical power, self-report measures were used for the manipulation check rather than the physiological measures. The PANAS was administered at three different points throughout the session: one prior to stress induction, one directly after stress induction, and one about 10 minutes after stress induction. The total negative affect score for each PANAS was captured for 75 participants. The descriptive statistics in Table 5 show that mean negative affect scores show the expected descriptive pattern based on the intended stress induction.

A multi-level linear model with a random intercept term was run on the negative affect PANAS data for the three stress groups at three time points. This accounted for individual differences in effect of the stress factor. Individual differences between participants accounted for 56% of the variance in negative affect observed. It also showed that negative affect was significantly lower at debrief than at pre-stress induction ($B=-0.81$, $p=0.022$), and that during stress induction, the high stress group was significantly more stressed than the control group ($B=1.5$, $p=0.009$), while the medium stress group was not significantly more stressed than the control group ($B=0.69$, $p=0.192$). This suggests that participants started and returned to similar stress levels, but that during stress induction they showed differences in their stress levels.

Table 4: Descriptive statistics for heart rate

Heart rate change	Group	Mean	SD	*n*
Baseline to stress induction	Control	3.75	0.22	3
	Moderate stress	4.38	1.02	3
	High stress	2.73	2.88	4
	Total	3.53	1.89	10
Stress induction to post-stress	Control	1.31	1.86	3
	Moderate stress	0.34	3.93	3
	High stress	1.23	4.78	4
	Total	0.99	3.47	10

Table 5: Descriptive statistics for self-report stress in Experiment 2 at three time points: pre-stress, stress induction and debrief

Time	Stress group	Mean	SD	*n*
Pre-stress	Low	5.29	1.88	31
	Moderate	4.80	2.10	25
	High	4.68	2.50	19
Stress induction	Low	5.16	1.88	31
	Moderate	5.36	2.10	25
	High	6.05	3.06	19
Debrief	Low	4.48	1.86	31
	Moderate	4.24	1.72	25
	High	4.05	1.78	19

A 3x3 multi-level linear model, for which the factors are the three levels of stress and three time points, showed a main effect for time ($F(2,144)=13.22$, $p<0.001$, $Eta^2=0.152$), but not for group ($F(2,72)=0.07$, $p=0.617$, $Eta^2=0.001$). The overall interaction between time x group approached significance ($F(4,148)=6.27$, $p<0.086$, $Eta^2=0.042$). This was investigated further, as ideally there would be a significant difference only at time point 2, when the stress was induced. A planned contrast of this point showed a non-significant trend of induced stress across the groups (B=0.89, SE=0.61, $p=0.144$) with the high stress group having greater negative affect than the low stress ($F(2,72)=1.68$, $p=0.199$) or moderate stress ($F(2,72)=0.82$, $p=0.370$) groups and the moderate stress group having greater negative affect than the low stress group ($F(2,72)=0.154$, $p=0.697$).

Line-up performance

Data from 75 participants were used to test the hypothesis that there would be differences in identification accuracy across the stress groups. Table 6 shows that the medium stress group had the best performance in both TA and TP line-ups.

Table 6: Descriptive statistics for Experiment 2 by group, line-up and perpetrator

	Group		
Line-up decision	**Control**	**Moderate stress**	**High stress**
Target (perpetrator)			
TP	(*n* = 6)	(*n* = 17)	(*n* = 12)
Correct identification	67% (4)	76% (13)	58% (7)
TA	(*n* = 25)	(*n* = 8)	(*n* = 7)
Correct rejection	40% (10)	88% (7)	43% (3)
Target (bystander)			
TP	(*n* = 25)	(*n* = 8)	(*n* = 7)
Correct identification	64 % (20)	87% (7)	71% (5)
TA	(*n* = 6)	(*n* = 17)	(*n* = 12)
Correct rejection	33% (2)	71% (12)	58% (7)
Target (both)			
TP	(*n* = 31)	(*n* = 25)	(*n* = 19)
Correct identification	64 (20)	80 (20)	63 (12)
TA	(*n* = 31)	(*n* = 25)	(*n* = 19)
Correct rejection	39% (12)	76% (19)	53 (10)

TP, target present; TA, target absent

This can be seen in the greater proportion of hits and lower proportion of false alarms for the medium stress group, as well as lower rates of rejection and 'don't know' responses in the TP line-ups, and higher rates of correct rejections in the TA line-ups. Logistic regressions of identification accuracy on stress were run using stress manipulation as a dummy variable, as well as a continuous variable, to assess whether effects would be linear or quadratic. As there was no significant difference between roles (B=0.25, SE=0.40, $p=0.527$) data were collapsed across perpetrator and bystander roles to increase statistical power.

The results of the regression analyses suggest a non-linear relationship. For the logistic regression using stress as a categorical predictor, one can see that the medium stress group always performed better than the low stress group (B=1.20, SE=0.43, $p=0.005$) and that there was no difference between the low and high groups (B=0.25, SE=0.42, $p=0.541$), as would be predicted from the negative quadratic relationship shown in Figure 2. Trend analyses showed that a negative quadratic model (B= -6.14, SE=2.28, $p=0.007$) was significant and the linear model was not (B=2.12, SE=2.02, $p=0.294$). A chi-squared model comparison showed that the quadratic model was significantly better than a linear one ($\chi^2 2(1, 147) = 7.86$, $p=0.005$). These two ways of looking at the data both confirm that the moderate stress group were the best performers, as expected under the Yerkes-Dodson relation. Table 7 shows the full sets of comparisons, where the non-linear models were consistently better than the linear ones.

Discussion

The experiments presented here investigated the effect of stress on eyewitness memory. Both used a live encoding event and a photographic line-up presented more than 24 hours later. Experiment 1 used a two-group (high/low) stress manipulation, with careful attention to methodological features to ensure successful stress induction, and found no significant effect of stress on witness memory, replicating Sauerland et al.[16] Experiment 2 used a three-level stress manipulation, inducing descriptively different stress levels, and found a non-linear relationship between stress intensity at encoding and line-up performance. While the power of the experiments is not sufficient to draw strong conclusions, the studies suggest that using only two levels of stress is not sufficient to show the effects of stress on witness memory. Although this is the current trend in the literature, a two-group design does not have the necessary resolution to reliably capture a non-linear relationship. Better methods of stress induction, using more than two groups, are absent in the literature and this absence needs to be addressed so as to avoid cumulation of null results or results of conflicting direction, due to low resolution modelling. Experiment 2, which found no difference between the low and high stress groups, confirms this, as without the moderate stress level, the pattern matches that of Experiment 1, in which there was no observable effect of stress at encoding. Figure 2 illustrates this problem, as depending on the level of induced stress, it is possible to find a positive, negative or null difference between stress groups (lines A and B). As inspection of the Yerkes-Dodson law[8] suggests, a non-linear relationship, measuring three or more points of stress, should become the standard practice.

Reasons for the non-linear relationship of stress and witness memory may be both physiological and behavioural. The SAM axis responses to stressors encourage the 'fight or flight' response, which may improve encoding: by focusing attention on the source of the stressor, it is more likely to be better remembered. A strong stress response will release hormones, which may impair encoding of witnessed events. These contrasting components should be better studied, and effects mapped so that the stress response can be viewed in terms of its separate physiological pathways. This could well improve the predictive utility of stress research, which extends to other research on memory as well as other physiological systems. Better manipulation checks using both physiological and standardised self-report measures will allow for clearer understanding of stress effects on witness memory. While this study showed a non-linear effect of stress intensity, there are still other factors which likely moderate the stress–performance relationship.

One of these factors might be the type of memory task utilised. In Experiment 1, there was a trend in the recall data that showed the low stress group performed slightly better than the high stress group – a trend which was not seen in the recognition data. This might indicate that

Table 7: Comparison of linear and non-linear models

Analysis	Condition	Comparison	Coefficient	SE	Z	p	AIC
Linear logistic regression	TP perpetrator	2 3	0.49 -0.36	1.04 1.05	0.47 -0.34	0.640 0.733	48.49
	TA perpetrator	2 3	2.35 0.12	1.14 0.87	2.01 0.14	0.040* 0.892	55.24
	TP bystander	2 3	1.37 0.34	1.15 0.93	1.20 0.37	0.232 0.715	53.08
	TA bystander	2 3	1.57 1.03	1.02 1.05	1.54 0.99	0.123 0.325	50.54
	TP both targets	2 3	0.79 -0.06	0.63 0.61	1.26 -0.10	0.207 0.923	96.35
	TA both targets	2 3	1.61 0.56	0.60 0.59	2.71 0.96	0.007* 0.337	101.22
	Collapsed TP and TA across targets	2 3	1.20 0.25	0.43 0.42	2.82 0.61	0.005* 0.541	196.3
Polynomial logistic regression	TP perpetrator	Linear Quadratic	-1.20 -1.91	2.08 2.22	-0.58 -0.86	0.564 0.392	48.49
	TA perpetrator	Linear Quadratic	1.98 -5.55	2.11 2.79	0.94 -1.99	0.349 0.047*	55.24
	TP bystander	Linear Quadratic	1.72 -2.90	2.23 2.82	0.77 -1.03	0.443 0.304	53.075
	TA bystander	Linear Quadratic	1.38 -3.03	2.08 2.14	0.66 -1.41	0.508 0.158	50.54
	TP both targets	Linear Quadratic	0.27 -3.31	2.08 2.36	0.13 -1.40	0.898 0.162	96.35
	TA both targets	Linear Quadratic	2.73 -5.37	2.03 2.24	1.34v-2.40	0.180 0.016*	101.22
	Collapsed TP and TA for both targets	Linear Quadratic	2.12 -6.14	2.02 2.28	1.05 -2.69	0.294 0.007*	196.3

TP, target present; TA, target absent; AIC, Akaike information criterion

tasks taxing different memory systems are affected differently by stress. However, it might also be an indication that task difficulty is an important moderating variable because the recall task was likely more difficult than the recognition task.[12] Although we included depressive mood as a covariate in our analyses, it was not found to predict any difference in performance, perhaps because the average level in our sample of students was low, and more severe depressive symptoms may be a moderating factor of the relationship between stress and memory.

Experiment 2 found no difference in recognition for the perpetrator who induced the stress and a bystander who appeared afterwards. This suggests that the effect of stress on facial encoding and recognition is generally as a result of the neurochemical response and not tied to the source of the stressor. While the stress induction for Experiment 2 did not show significant differences, it did identify the expected trend using both physiological and self-report measures. The stress induced was an inherent part of the eyewitness event, rather than induced through a laboratory manipulation as in Experiment 1. While such methods may not be as robust, they may be more meaningful to the participants. As stress is both subjective and physiological, the meaning and consequences of the stress may differ depending on its source. As such, the intensity, source and meaning of the induced stress should be considered when investigating the effect of stress on witness memory.

In conclusion, the studies presented here suggest that the stress–performance relationship should be considered non-linear. The main limitation in both of these studies is the relatively low sample size and consequent low statistical power. As such, we note that the studies are not conclusive but rather that the descriptive results suggest that the methods used for stress induction in the literature should be reconsidered. Experiments aiming to explore the effect of stress on eyewitness memory for faces and details of crimes should make use of designs with at least three stress groups. This will allow future studies to better investigate the stress–performance relationship along with any moderating effects.

Competing interests

We have no competing interests to declare.

Authors' contributions

M.G.: Conceptualisation, methodology, data collection, data analysis, data curation, writing – the initial draft, writing – revisions, project management. T.J.: Conceptualisation, methodology, data collection, writing – revisions. C.T.: Conceptualisation, methodology, writing – revisions, student supervision, project leadership.

References

1. Deffenbacher KA, Bornstein BH, Penrod SD, McGorty EK. A meta-analytic review of the effects of high stress on eyewitness memory. Law Hum Behav. 2004;28(6):687–706. https://doi.org/10.1007/s10979-004-0565-x

2. Gering MA, A systematic review and meta-analysis on the effects of stress at encoding on line-up decisions [master's thesis]. Cape Town: University of Cape Town; 2021.

3. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol. 1908;18(5):459–482. https://doi.org/10.1002/cne.920180503

4. Wolf OT, Atsak P, De Quervain DJ, Roozendaal B, Wingenfeld K. Stress and memory: A selective review on recent developments in the understanding of stress hormone effects on memory and their clinical relevance. J Neuroendocrinol. 2016;28(8):1–8. https://doi.org/10.1111/jne.12353

5. Gray J, Kogan J. Marrocco, McEwen, B. Genomic and epigenomic mechanisms of glucocorticoids in the brain. Nat Rev Endocrinol. 2017;13(11):661–673. https://doi.org/10.1038/nrendo.2017.97

6. Shields GS, Sazma MA, Yonelinas AP. The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. Neurosci Biobehav Rev. 2016;68(1):651–668. https://doi.org/10.1016/j.neubiorev.2016.06.038

7. Roozendaal B, McEwen BS, Chattarji S. Stress, memory and the amygdala. Nat Rev Neurosci. 2009;10(6):423–433. https://doi.org/10.1038/nrn2651

8. Drexler SM, Wolf OT. Stress and memory consolidation. In: Axmacher N, Rasch B, editors. Cognitive neuroscience of memory consolidation. Studies in Neuroscience, Psychology and Behavioral Economics. New York: Springer; 2017 p. 285–300. https://doi.org/10.1007/978-3-319-45066-7_17

9. Sandi C. Stress and cognition. Wiley Interdiscip Rev Cogn Sci. 2013;4(3):245–261. https://doi.org/10.1002/wcs.1222

10. Levine LJ, Edelstein RS. Emotion and memory narrowing: A review and goal-relevance approach. Cogn Emot. 2009;23(5):833–875. https://doi.org/10.1080/02699930902738863

11. Vakil E, Wasserman A, Tibon R. Development of perceptual and conceptual memory in explicit and implicit memory systems. J Appl Dev Psychol. 2018;57(1):16–23. https://doi.org/10.1016/j.appdev.2018.04.003

12. Het S, Ramlow G, Wolf OT. A meta-analytic review of the effects of acute cortisol administration on human memory. Psychoneuroendocrinology. 2005;30(8):771–784. https://doi.org/10.1016/j.psyneuen.2005.03.005

13. Gotlib IH, Joormann J. Cognition and depression: Current status and future directions. Annu Rev Clin Psychol. 2010;27(6):285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

14. Bearden CE, Glahn DC, Monkul ES, Barrett J, Najt P, Villarreal V, et al. Patterns of memory impairment in bipolar disorder and unipolar major depression. Psychiatry Res. 2006;142(3):139–150. https://doi.org/10.1016/j.psychres.2005.08.010

15. Wells GL, Kovera MB, Douglass AB, Brewer N, Meissner CA, Wixted JT. Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. Law Hum Behav. 2020;44(1):3–36. https://doi.org/10.1037/lhb0000359

16. Sauerland M, Raymaekers LH, Otgaar H, Memon A, Waltjen TT, Nivo M, et al. Stress, stress-induced cortisol responses, and eyewitness identification performance. Behav Sci Law. 2016;34(4):580–594. https://doi.org/10.1002/bsl.2249

17. Smeets T, Cornelisse S, Quaedflieg CW, Meyer T, Jelicic M, Merckelbach H. Introducing the Maastricht Acute Stress Test (MAST): A quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. Psychoneuroendocrinology. 2012;37(12):1998–2008. https://doi.org/10.1016/j.psyneuen.2012.04.012

18. Gabbert F, Hope L, Fisher RP. Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. Law Hum Behav, 2009;33(4)298–307.

19. Team RC. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/

20. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48. https://doi.org/10.18637/jss.v067.i01

21. Diamond DM, Campbell AM, Park CR, Halonen J, Zoladz PR. The temporal dynamics model of emotional memory processing: A synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. Neural Plas. 2007;2007(1):1–33. https://doi.org/10.1155/2007/60803

22. Baritaki S, De Bree E, Chatzaki E, Pothoulakis C. Chronic stress, inflammation, and colon cancer: A CRH system-driven molecular crosstalk. J Clin Med. 2019;8(10):1669–1695. https://doi.org/10.3390/jcm8101669