Received: 17 November 2020 | Revised: 15 December 2021 | Accepted: 16 December 2021

DOI: 10.1002/acp.3910

RESEARCH ARTICLE

WILEY

Effects of stress on eyewitness identification in the laboratory

Heather L. Price[1] | Laurie Sykes Tottenham[2] | Bianca Hatin[3] | Ryan J. Fitzgerald[4] | Eva Rubínová[1]

[1]Department of Psychology, Thompson Rivers University, Kamloops, British Columbia, Canada

[2]Department of Psychology, University of Regina, Regina, Saskatchewan, Canada

[3]Department of Psychology, University of the West of Scotland, Paisley, UK

[4]Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada

Correspondence
Heather L. Price, Department of Psychology, Thompson Rivers University, 805 TRU Way, Kamloops, BC V2C 0C8C, Canada.
Email: hprice@tru.ca

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: 2017-04986; University of Regina Dean's Research Award; Canada Foundation for Innovation

Abstract

Witnesses to crime often experience stress during the witnessed event. However, most laboratory studies examining eyewitness memory do not include a stressful encoding event. Participants ($N = 129$) completed an experimental stress induction procedure—a modified version of the Trier Social Stress Test. We designed three conditions to manipulate the amount of stress experienced and included three types of measures to assess the effectiveness of the manipulation: cortisol levels (hormonal), blood pressure and heart rate (autonomic), and self-report (subjective). Participants watched a video that had a surprise viewing of a staged theft and completed two lineup identification tasks. We observed no effects of stress on the accuracy or willingness to choose from a lineup. Importantly, there was variability in the correspondence between measured indicators of stress, which should be considered in future designs.

KEYWORDS

encoding, eyewitness, identification, laboratory, stress

1 | INTRODUCTION

Witnesses to crime often experience stress during the witnessed event. However, most laboratory studies examining eyewitness memory do not include a stressful encoding event. Neglect of a stress manipulation matters little if stress at encoding is unrelated to identification accuracy, but if there is a relationship, then awareness of encoding stress will influence estimates of identification accuracy. Given that much of the literature from which procedural recommendations are derived involves encoding events that do not inherently elicit high levels of emotion (e.g., watching innocuous mock crime videos), it is critical to better understand the impact of stress on eyewitness identification.

From the extant literature, it is unclear how witness stress at the time of the offense affects subsequent accuracy at a lineup identification procedure. The difficulty of understanding the effects of stress likely arises from at least one of several possibilities. Perhaps most critically, ecologically valid inductions of stress are difficult to achieve ethically. Further, measurements of stress in the existing literature have been inconsistent and have largely relied on subjective judgments. Finally, there is substantive methodological variability across studies. Thus, drawing conclusions across studies is difficult. In the present research, we sought to induce three levels of stress in adult eyewitnesses using a modified version of an in-person standardized task (Trier Social Stress Test; TSST; Kirschbaum et al., 1993) and measured our manipulations of stress using both subjective (self-report) and objective (autonomic and hormonal) methods. The aim of the present work was to add to the slowly growing body of literature on adult eyewitness identification after a stressful event by obtaining multiple measures of stress and using stress induction tasks based on well-established and validated approaches.

A meta-analysis on the effects of stress on eyewitness identification accuracy was conducted more than 15 years ago by Deffenbacher and colleagues (Deffenbacher et al., 2004). The meta-analysis explored the influence of encoding stress in 22 eyewitness identification studies, of which 15 involved adult witnesses and only seven reported physiological measures of stress. The authors generally concluded that stress during encoding negatively impacts target-present (but not target-absent) identification performance. However, conclusions were tentative due to the substantial variability in methodologies and small number of available studies. Recent evidence suggests that eyewitness experts continue to struggle to provide a conclusive statement about the effects of encoding stress on identification accuracy, and that the conclusions of eyewitness experts differ from that of fundamental memory experts who employ less applied paradigms (Marr, Otgaar, et al., 2021).

Since the Deffenbacher et al. (2004) review was published, only a few studies have examined the relation between encoding stress and identification performance in adults, and only three have included physiological measures (see Rush et al., 2014, e.g., with child eyewitnesses). The inclusion of physiological measures is important because subjective and physiological measures of stress are often unrelated (Hellhammer & Schubert, 2012; but see Valentine & Mesout, 2009). There are, however, some impressive exceptions to this general dearth of literature.

In one example of a particularly effective stress induction, Morgan et al. (2004) tested soldiers' identification accuracy (after a 24-h delay) of an interrogator encountered in a mock prisoner of war camp. The soldiers were held for 12 h and interrogated with either physical confrontation (high stress) or without physical confrontation (low stress) for more than 30 min. Across four studies with over 500 participants, Morgan et al. reported a general pattern that high stress impaired identification accuracy. Hope et al. (2012) reported similar results in a study comparing memory of law-enforcement officers who underwent a physically exerting assault scenario and (control) officers who observed this scenario. Physical exertion (validated by an increase in heart rate) was associated with lower target identification and higher filler identification rates in a target-present lineup (a target-absent lineup was not administered).

In another creative study, Valentine and Mesout (2009) assessed witnesses' memories for a target encountered in the Horror Labyrinth of the London Dungeon. The authors initially ran a sample of 18 test participants through the Labyrinth to assess the relationship between heart rate and a state anxiety measure, and found a significant correlation between the state measure and heart rate. Then, the full sample of 56 participants completed the state anxiety measure after visiting the Dungeon and encountering the target. Using target-present lineups only, the authors found that 17% of participants who scored above the median on the state anxiety scale correctly identified the target, whereas 75% of participants who scored below the median correctly identified the target. Thus, Valentine and Mesout also found a negative effect of high levels of stress on identification accuracy.

Most recently, Sauerland et al. (2016) explored the apparent contradiction between the (relatively weak) conclusions drawn from the eyewitness identification literature that stress at encoding negatively impacts identification accuracy (Deffenbacher et al., 2004) and the cognitive neuroscience literature that shows an inverted U-shaped relationship, characterized by higher performance after moderate encoding stress than after high or low stress (see LaBar & Cabeza, 2006). Sauerland et al. (2016) used the Maastricht Acute Stress Test (MAST; Smeets et al., 2012) in which participants immersed their hand in ice water (or a control lukewarm bath) according to a schedule and completed mental math calculations (or a simple control counting task), thus creating stress and no stress conditions. During a break in the procedure, participants were exposed to an in-lab staged theft of a phone. One week later, 123 participants returned and completed a simultaneous photo lineup. The authors concluded there was no difference in identification accuracy between participants who did and did not experience stress at encoding after a one-week delay.

These studies involve creative methodology and each has considerable strengths, but each also has limitations. For example, in the Morgan et al. (2004) studies, participants were assigned to high or low stress interrogations (or both), but there were no measures of stress to assess the efficacy of the manipulation. In the Hope et al. (2012) work, participants were physically stressed, but there was no induction of mental stress. In the Valentine and Mesout (2009) work, a small sample size and a lack of physiological measures on the full sample limit potential conclusions. For the most recent Sauerland et al. (2016) paper, many limitations of prior research were addressed, but only a single measure of stress was assessed (cortisol). Further, all studies dichotomized stress (high/low or present/absent) and the methodologies used to elicit stress reactions differ so substantially across studies that drawing common conclusions is challenging.

2 | THE PRESENT STUDY

Here, we add to the dearth of literature examining stress and eyewitness identification accuracy by examining stress using multiple measures. A recent review by Shields et al. (2017) indicated that as the stressor-encoding delay increases, stress has an increasingly negative effect on encoding, with impairment taking place at 11 min post-stressor (with a significant impairing effect after 22 min). Further, Shields et al. found that this effect was moderated by relations between the stressor and the encoded information, with an impairment effect observed when a stressor was unrelated to the memory task. In the present work, we administered an experimental procedure in which stress was induced a few minutes prior to encoding and continued post-encoding; thus, although the initial induction of stress was not 11 min prior to the encoding task, the duration of stress experienced was. Further, our identification task was unrelated to the stressor, another condition that would predict a detrimental effect of stress on accuracy. Thus, our design leads to a prediction of a negative impact of stress on recognition.

We administered two identification tasks and collected several measures of stress from a final sample of 129 adults, who signed up for a study on stress in an employment context, experienced a stressful event, and had a surprise viewing of a staged crime video. The procedure was administered in individual sessions which lasted approximately 1.5 h. We designed three conditions to manipulate the amount of stress experienced and included four measures to assess the effectiveness of the manipulation: cortisol levels (hormonal), blood pressure and heart rate (autonomic), and self-report (subjective). To assess relations between task difficulty and stress, lineups comprised fillers that were either higher or lower in similarity to the target. Research has shown that high filler similarity impedes target identification (Clark, 2012; Fitzgerald et al., 2013; Oriet & Fitzgerald, 2018), and that higher task difficulty is generally associated with increased perceived stress and increased sympathetic activity, including increased blood pressure (Britt, 2005; Callister et al., 1992). We anticipated that filler similarity might interact with stress in that more challenging lineups (i.e., higher filler similarity) might combine with higher levels of stress to contribute to lower identification accuracy, relative to conditions of lower stress and difficulty. Finally, we expected that we might observe individual differences in participants' evinced stress responses across each of the measures, but we did not develop specific hypotheses about the nature of those differences.

3 | METHOD

3.1 | Participants and design

We employed a mixed 3 (stress: low vs. moderate vs. high) $\times$ 2 (target-presence: present vs. absent) $\times$ 2 (lineup similarity: lower vs. higher) design. Stress was a between-subjects variable, whereas target-presence and similarity were within-subjects variables. Participants completed two lineups, which were associated with different actors from the staged crime video (one for the person who acted as the thief and another for the person who acted as a research participant). The dependent variables were identification responses and post-identification confidence ratings. All participants were recruited through a departmental research participant pool and compensated with course credit. This project was approved by the university's Research Ethics Board.

3.1.1 | Eligibility criteria

Participants were required to be between 18 and 40 years of age. They were not eligible if they had taken hormone supplements in the past 3 months (other than oral contraceptives) or were on other medications affecting cortisol or stress levels (including synthetic corticosteroids, beta-adrenergic antagonists, antidepressants, and amphetamines), if they had a hormonal illness or disease, or if they had a severe chronic medical condition. These exclusion criteria were implemented to avoid factors known to be associated with cortisol concentrations (e.g., Alexander et al., 2007; Kudielka, Buske-Kirschbaum, et al., 2004a; Kudielka, Buske-Kirschbaum, et al., 2004b; Kudielka, Schommer, et al., 2004; Paananen et al., 2015; Wand et al., 2007). Further, students were not eligible if they had an oral injury or disease that could cause bleeding or high bacteria, which would affect the accuracy of the salivary hormone assays. Students were also ineligible if they did not follow the instructions for sleeping, or did not avoid exercise and food/drink consumption before testing (see screening measure below).

3.1.2 | Participants

The final sample included 129 participants ($M_{age} = 20.45$, $SD = 2.86$; range 18–34 years; $n = 104$ females; 55% Caucasian). An additional 44 participants were excluded for taking contraindicated medications ($n = 9$), not adhering to the saliva collection preparation instructions described below ($n = 7$), consuming caffeine within 3 h of being tested ($n = 4$), eating a meal within an hour of being tested ($n = 3$), or other reasons (1 withdrew data after participation, 1 had wisdom teeth removed 2.5 weeks earlier, 9 dropped-out due to discomfort with the saliva sampling or stress manipulation, 3 had language comprehension difficulties, and 7 had BMI scores ≥30).

Of the female participants, 52 took hormonal contraceptives and 52 were naturally cycling. Of the naturally cycling women, 26 were tested in the follicular phase and 23 were tested in the luteal phase (based on self-report of next and previous period and 28-day cycle); 3 did not report phase information. No participants were habitual smokers and none were pregnant.

3.2 | Materials and measures

3.2.1 | Blood pressure and heart rate monitor

Blood pressure (systolic and diastolic: BPS and BPD) and heart rate (HR) measures were taken via a cuff that connected wirelessly to the recording device. The cuff was affixed to the participant's wrist on their nondominant hand, which was supported by a pillow to provide stability and the required elevation. Readings were taken at 12 points during the experiment, during seven different periods (multiple readings were taken for longer tasks, see Figure 1). Due to wireless connectivity issues, there were some missing values in the BP/HR measures (for 6 participants no BP/HR data were available, and for 12 participants data were not available for one or more periods). Due to violations of the normality assumption in the distributions of the BPS and HR measurements, we: (i) excluded 12 BPS values (<70 and >160) that were identified as outliers to reduce high kurtosis, and (ii) log-transformed HR values to reduce high skewness and kurtosis.

3.2.2 | State–Trait anxiety inventory

The state anxiety scale of the State–Trait Anxiety Inventory (STAI; Spielberger et al., 1983) was administered three times throughout the session. This 20-item scale, designed to assess current emotional anxiety, has shown good internal consistency alpha coefficients (0.86–0.95; Spielberger et al., 1983), and is recommended for use with the TSST (e.g., Birkett, 2011; Narvaez Linares et al., 2020). The STAI-trait was administered once at the start of the session to measure general anxiety levels. This 20-item scale has also shown good internal consistency alpha coefficients (0.89–0.90; Spielberger et al., 1983).

3.2.3 | Saliva sample, assay, and screening measure

To control factors that could influence cortisol concentrations or the accuracy of the saliva assay, the recruitment message informed

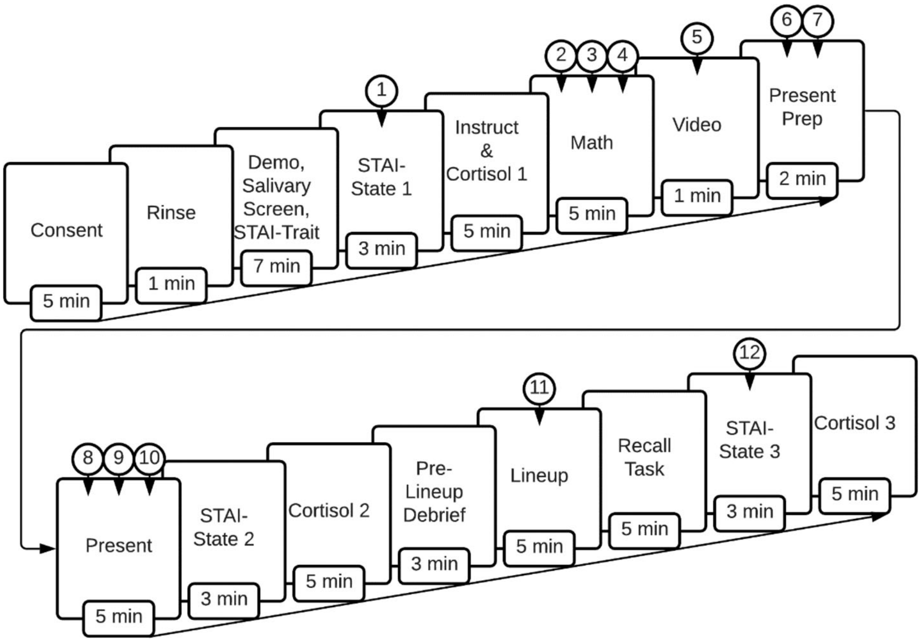

FIGURE 1 Experimental procedure. Circles indicate heart rate and blood pressure measurements

students that only those who adhered to the following requirements would be permitted to participate: (1) no alcohol consumption within 12 h or vigorous physical activities within 90 min of the session; (2) no caffeine use within 3 hours of the session; (3) nothing in the mouth except for water within 1 h of the session (e.g., no food/gum/dental floss/beverages other than water); and (4) wake at least 3 h prior to the session. Adherence to these instructions was assessed using a self-report screening measure administered at the testing session. Participants were tested between 11:00 am and 4:30 pm[1] to limit the influence of diurnal cortisol fluctuations.

Participants provided three saliva samples (approximately 1 mL each) during the testing session: at baseline, after the 2nd STAI-state which was immediately following the experimental stress induction procedure, and at the end of the testing session (Figure 1). Following collection, the samples were immediately frozen at -40°C and stored until data collection was complete. The samples were subjected to enzyme immunoassay (EIA), using a commercially available kit (Salimetrics), in order to determine salivary cortisol concentrations. The EIA analyses were performed in the SPIT laboratory at the University of Regina. All samples were assayed in duplicate. The EIA procedure produced intra- and inter-assay coefficient of variations of 6.62% and 8.63%, respectively. The values of cortisol measurements were log-transformed to reduce high skewness and kurtosis.

3.2.4 | Video

The target event was a video (70 s in duration) that involved three male actors in the roles of research participant, victim, and thief. Six versions of the video were produced to counterbalance the actors across roles. The video began with the research participant reading a speech about his strengths and weaknesses. Part-way through the speech, the camera shifted focus to a student (victim) sitting on a chair in the background. The student received a phone call and exited the scene, leaving his laptop bag behind. The thief subsequently entered the scene, rummaged through the bag, and exited with a laptop, at which point the camera focus shifted back to the participant. The research participant and thief (i.e., the two video characters who served as targets for the identification tasks) were each in view for 20 s.

3.2.5 | Lineups

Judges ($n = 8$–9) rated the similarity between each of the three actors and 72–75 potential fillers. All potential fillers were matched to the actors' sex, race, and age. Ratings were made on a 10-point Likert scale. The scale was subdivided into the categories of high dissimilarity (1–3), moderate similarity (4–7), and high similarity (8–10). Mean similarity ratings indicated that judges considered all potential fillers to be either highly dissimilar or moderately similar. For each actor, we used these ratings to construct lineups that differed in target-filler similarity (see Table 1). We refer to lineups containing fillers from the highly dissimilar and moderately similar categories as "lower" and "higher" similarity lineups, respectively.

In addition to target-filler similarity, we manipulated whether or not the target actor was in the lineup. The target-present lineups included an actor and five fillers. Target-absent lineups were created by taking the target-present lineups and replacing the target with an "innocent suspect" who resembled the target actor but was not in the video. Witnesses viewed the lineup members simultaneously, in a 2×3 array. The spatial location of lineup members was counterbalanced across all six positions.

TABLE 1 Lineup filler similarity ratings

	Entire set		Lower similarity lineup		Higher similarity lineup		Innocent suspect	
Actor	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
A	3.88	0.91	3.18	0.19	5.28	0.30	6.00	1.07
B	4.29	0.85	3.16	0.30	5.29	0.27	5.89	2.15
C	3.96	0.88	3.18	0.17	5.18	0.42	5.78	1.92

3.2.6 | Demographics questionnaire

A demographics questionnaire was administered to assess factors that could have affected cortisol concentrations or task performance. Factors included: age; sex; handedness; race; height; and weight (for calculation of body mass index [BMI]); medication, steroid hormone, and nicotine use; vision and hearing problems; history of hormonal, neurological, and psychiatric illnesses; and chronic and acute medical conditions. Female participants were also asked to report on their menstrual cycle regularity and phase, and whether or not they were pregnant.

3.3 | Procedure

Undergraduate students signed up for a study entitled, "Stress Effects on Job Preparation Skills." The recruitment message explained that the purpose of the study was to understand the physiological responses associated with job interviews and that participation would involve watching a video, completing job applicant tasks and questionnaires, and having physiological and hormonal measures taken.

Upon arrival, students reviewed and signed a consent form. Next, participants rinsed their mouths to remove potential contaminants in preparation for collection of the saliva samples. After the rinse, the experimenter fastened the heart-rate monitor to participants' wrists and administered three questionnaires: demographic, salivary screen, and STAI (state and trait), followed by the first BP/HR reading. Participants then provided the first saliva sample, approximately 10 min after rinsing.

The experimental stress induction procedure was a modified version of the TSST (Kirschbaum et al., 1993). While a number of modifications were needed in order to manipulate stress level and include a target event for the purposes of the present study, efforts were made to adhere to recent recommendations (Birkett, 2011; Narvaez Linares et al., 2020) for the administration of the TSST with respect to the timing and use of subjective, autonomic, and endocrine stress measures. This procedure began with a 5-min computerized mathematics exercise, with responses being provided with the dominant hand via keyboard. After the math exercise, participants watched the target video under the pretense that they would be watching an example of a former participant giving a speech on his strengths and weaknesses, as an illustration of what was expected of them for their speeches. Immediately after the video, participants were given 2 min to prepare their speeches, and then delivered their speeches for a duration of 5 min. Thus, participants were exposed to stress immediately before and after the target video. BP/HR readings were taken during the math exercise (3), video (1), speech prep (2), and speech (3). The difficulty of the math exercise and the social-evaluative aspect of the speech-related procedures depended on whether the participant was in the low, moderate, or high stress condition (manipulation procedures detailed below).

Immediately following the speech, participants completed a second STAI-state and provided a second saliva sample (approximately 15 min after the start of the TSST). Participants were subsequently informed that the true purpose of the study was to examine eyewitness memory and asked if they were willing to proceed with some memory tasks. Those who consented were informed that they would complete two lineup identification tasks, one for the thief and one for the research participant in the video (order counterbalanced). The 6-member lineups were administered on a computer by a blind administrator, during which a BP/HR reading was taken. Half of the lineups included the target and the other half did not. Lineups were also manipulated in terms of target-filler similarity. Prior to the identification tasks, participants were warned that the actor may or may not be in the lineup and that they were not required to choose any of the lineup members. A BP/HR reading was taken during the identification task. After making an identification decision, participants were asked to indicate their identification confidence using a 6-point Likert scale.

Following the identification task, participants were also asked if they noticed anything unusual about the photos in the lineup, if they had participated in an eyewitness identification study previously, if they were or became aware of the study purpose prior to it being disclosed, and if there was any reason their data should not be included in analyses. No participants were excluded based on their answers to these questions.

After the interview, participants completed a third STAI-state, had a final BP/HR reading taken, and provided a third saliva sample (approximately 35 min after the start of the TSST). Participants were then informed that the experimental procedures were complete. In the debriefing session, the experimenter asked participants to not share the nature of the study with their classmates, friends, or anyone else who could potentially participate in the study.

3.4 | Stress manipulation

3.4.1 | Low stress condition ($n = 48$)

Participants were given relatively easy math problems (addition or subtraction of 2-digit numbers, with values ranging between 10 and 99) and had unlimited time to complete each question in private. After the math test and target video, participants wrote a speech on their strengths and weaknesses as a job candidate. They subsequently gave the speech in private, spoken quietly to themselves.

3.4.2 | Moderate stress condition ($n = 43$)

Participants completed more difficult math questions (addition or subtraction of 3-digit numbers, with values ranging between 100 and 999), during which the administrator remained in the room. Participants were required to answer each question within a 20 s time limit. Immediately after each response was entered, a screen appeared for 2 s indicating whether it was correct or incorrect; an alarm sound also played following incorrect responses. After the math test and target video, participants prepared and presented a speech on their strengths and weaknesses as a job candidate. Participants were instructed that the presentation would be videotaped for the purposes of record-keeping, and the administrator remained in the room. The video recorder was not actually turned on.

3.4.3 | High stress condition ($n = 38$)

Participants completed math questions with the administrator in the room. The questions were the same difficulty as in the moderate stress condition, but participants were given only 7 s to provide each response. Again, a screen appeared for 2 s immediately after each response indicating whether it was correct or incorrect and an alarm sound played following incorrect responses. Consistent with the moderate stress condition, participants prepared and presented a speech on their strengths and weaknesses as a job candidate in front of the administrator. However, participants in the high stress condition were told that they would be videotaped and the recording would be shown to future participants who would critique their performance. The video recorder was not actually turned on.

3.5 | Statistical analyses

Due to the repeated nature of the physiological and self-report measurements in this study, and the repeated lineup decisions, we needed to account for the violation of independence in our data. We used linear mixed models (LMMs) that allowed (where applicable) intercepts and slopes to vary across participants, measurements, and lineups (for lineups, actor identities were crossed with the roles they assumed in the video). In other words, the LMMs were used to apply a hierarchical structure to the analysis, where stress conditions, stress measurements, target presence, choosing, and similarity were at Level 2 and participants and lineup characters/actor were at Level 1. While accounting for this (random) source of variance, the models retained the statistical power of the sample to estimate the (fixed) effects of manipulated and measured stress, target presence and similarity, and choosing on eyewitness identification measures. Further model specifications are provided in the Results.

We performed grand-mean centering on all continuous variables (Enders & Tofighi, 2007) using the center function of the `misty` package (Yanagida, 2020). Categorical variables were coded using successive difference contrasts (from the package `MASS`; Venables & Ripley, 2002; Schad et al., 2020) to compare neighboring levels of multi-level factors. Therefore, results of the LMMs show main effect contrasts for each comparison of two levels of a categorical variable, for example, the difference in cortisol response between the low and moderate stress conditions and the moderate and high stress conditions.

The resulting regression coefficients are reported along with 95% confidence intervals (CI, shown in brackets) to show the range of their plausible values. In models with continuous dependent variables, regression coefficients for categorical independent variables reflect the mean differences between compared groups; regression coefficients for continuous independent variables indicate a degree of change associated with an increase in the independent variable. In models with binomial dependent variables, regression coefficients were converted to odds ratios (*OR*).

We used the `lme4` (Bates et al., 2015) and `lmerTest` (Kuznetsova et al., 2017) packages in R (R Core Team, 2020) to conduct statistical analyses, and the package `ggplot2` (Wickham, 2016) to produce figures. The data and R script can be found at: https://osf.io/37mp2/?view_only=94885577a96a4e7280434ce36acb929e.

4 | RESULTS

4.1 | Manipulation check

First, we explored the effectiveness of our stress manipulation within each of the three primary measures: autonomic response, cortisol, and subjective state anxiety.

4.1.1 | Autonomic response measures

We examined changes in BPS, BPD, and HR in the course of the stress-inducing procedure across the stress conditions. We did not include the last two measurements, which were collected during and after the lineup, as these were not stress-inducing portions of the experimental procedure (and the measurements indicated a drop in stress levels; see Figure 2 for BPS and Figure SM1 for BPD and HR data in Online Supplemental Materials). All three LMMs were defined with fixed effects of condition and timing of measurement (treated as a continuous variable), and with random intercepts for participant identity and random slopes for timing of measurement (the HR model with random slopes did not converge so measurement was defined as a random intercept effect).

We assessed within-subjects correlations for the repeated measurements using the `rmcorr` package (Bakdash & Marusich, 2017). All three autonomic stress indicators were significantly correlated, with the strongest correlation between the two indicators of blood pressure, BPS and BPD, $r = .66$ [.63, .69], $p < .001$; BPS and HR: $r = .28$ [.23, .34], $p < .001$; and BPD and HR: $r = .33$ [.27, .38], $p < .001$. Models for each

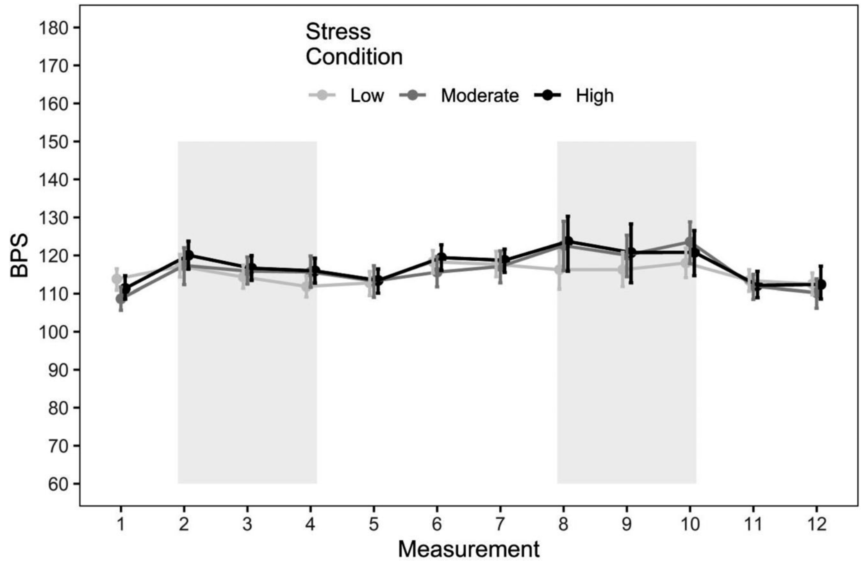

FIGURE 2 Systolic blood pressure across the course of the study in the low, moderate, and high stress conditions. Error bars represent 95% *CI*s of the means, and limits on the y-axis were set for the minimum and maximum measured values (raw data). Shaded areas represent stressful portions of the procedure (Mathematical task: measurements 2–4; Presentation: measurements 8–10). Reading 11 was taken during the lineup task

autonomic response measure revealed similar findings: there was a main effect of measurement timing showing that autonomic response increased across the course of the study, BPS: $\beta = .07$ [.05, .09], $t(118.96) = 7.19$, $p < .001$; BPD: $\beta = .09$ [0.07, 0.11], $t(120.05) = 8.61$, $p < .001$; and HR: $\beta = .14$ [.08, .19], $t(8.00) = 5.04$, $p = .001$ (complete statistics of all models can be found in Table SM1 in Online Supplemental Materials). There were no significant main effects of stress condition, *t*s < 1.60, *p*s > .115. For BPS, a significant interaction between condition and measurement indicated that the increase in BPS was greater for participants in the moderate stress condition than in the low stress condition, $\beta = .06$ [.02, .11], $t(118.25) = 2.66$, $p = .009$. The patterns of data in Figure 2 suggest that autonomic response was similar in the moderate and high stress conditions.

4.1.2 | Cortisol

The cortisol model was defined in a LMM with fixed effects of stress condition and timing of measurement and random intercepts for participant identity. The results revealed no significant main effects, *t*s < 1.83, *p*s > .069, and one significant interaction indicating a difference in the salivary cortisol response between the low and moderate stress conditions between the first and the second measurement, $\beta = .36$ [.10, .62], $t(252.00) = 2.69$, $p = .008$. Panel a of Figure 3 illustrates this interaction: as participants proceeded through the experiment, cortisol slightly increased in the moderate (and the high) stress condition but decreased in the low stress condition.

4.1.3 | STAI

The STAI model was defined with fixed effects of stress condition and timing of measurement (including the first two measurements that occurred during the stress-inducing portions of the procedure) and random intercepts for participant identity. There was a significant effect of condition: participants in the moderate stress condition indicated higher levels of subjective anxiety than participants in the low stress condition, $\beta = .35$ [.06, .64], $t(121.95) = 2.35$, $p = .020$. The contrast comparing the moderate and high stress conditions was not significant, $t(122.92) = .63$, $p = .533$. A significant effect of measurement timing reflected the pattern visible in panel b of Figure 3: subjective anxiety increased between the first two measurements, $\beta = 1.06$ [.92, 1.21], $t(119.15) = 14.24$, $p < .001$. A significant interaction between condition and measurement timing further specified that the increase was greater in the moderate than in the low stress condition: $\beta = 1.03$ [.69, 1.38], $t(118.48) = 5.85$, $p < .001$.

4.1.4 | Autonomic stress response, cortisol, and STAI

To assess the correspondence between subjective anxiety and cortisol levels and reflect the delay in the cortisol response, we computed peak differences from baseline (i.e., for STAI, we subtracted the first from the second measurement, and for cortisol, we subtracted the first from the third measurement). We needed to exclude cortisol values from seven participants to reduce kurtosis. We found a weak-to-moderate and significant correlation between subjective anxiety and cortisol levels, $r = 0.20$ [0.01, 0.37], $p = .036$.

To assess correlations between the autonomic stress measures and cortisol, we computed peak differences for BPS, BPD, and HR by subtracting the first from the tenth measurement (end of the presentation task). We needed to exclude BPS values from three participants and HR values from three participants to reduce kurtosis. We found weak-to-moderate and significant correlations between peak values for cortisol and BPS, $r = .24$ [.05, .41], $p = .014$, and also for cortisol and BPD, $r = .25$ [.06, .41], $p = .009$, but the correlation between cortisol and HR was not significant, $r = -.04$ [−.23, .15], $p = .671$.

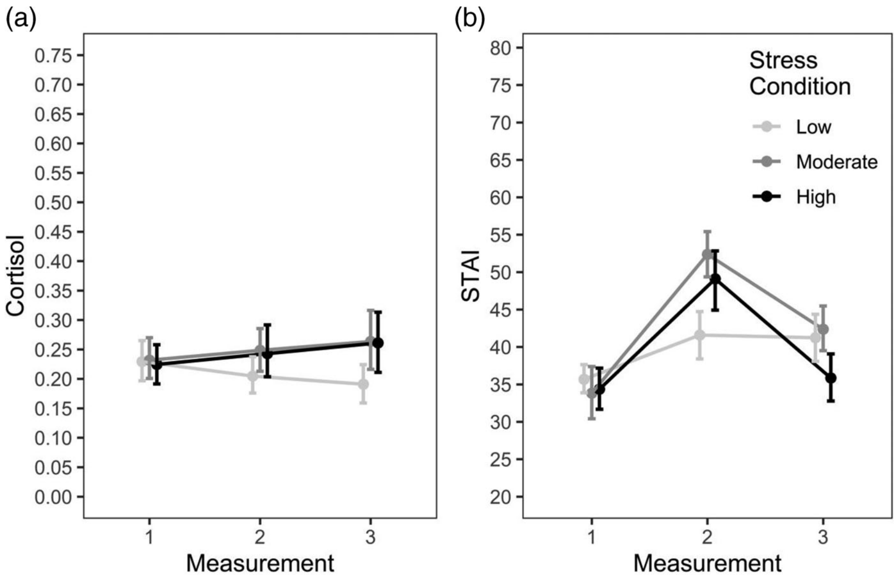

FIGURE 3 Salivary cortisol levels (non-transformed; Panel a) and subjective ratings of anxiety (Panel b) across the course of the study in the low, moderate, and high stress conditions. Error bars represent 95% CIs of the means, and limits on the y-axis were set for the minimum and maximum measured values (raw data)

To assess correlations between the autonomic stress measures and STAI, we selected BPS, BPD, and HR measurements 1, 10, and 12, which were taken at or near the time of the three STAI measurements (see Figure 1). Correlations between STAI and autonomic stress indicators were moderate-to-large and significant: BPS, $r = .48$ [.37, .57], $p < .001$; BPD, $r = .52$ [.42, .61], $p < .001$; and HR, $r = .44$ [.33, .54], $p < .001$.

4.1.5 | Manipulation effectiveness

Measures of autonomic, hormonal, and subjective responses indicated an increase in stress over the course of the study, with a stronger stress response in the moderate (and high) stress condition than in the low stress condition. Autonomic and subjective indicators then showed a return to baseline at the end of the procedure; the final hormonal stress measurement remained high (consistent with the typical delay and prolonged nature of the cortisol response). For all measures, differences were only found between the low and moderate stress conditions; responses in the moderate and high stress conditions were similar, suggesting that our high stress procedure did not produce a stress response exceeding that observed in the moderate stress condition. Therefore, in addition to conditions that resulted from the stress manipulation (i.e., our planned analyses), we explored the association between performance and the observed stress response as measured by autonomic, hormonal, and subjective indicators.

4.2 | Accuracy

Below, we provide results of multi-level regression models that, in general, explored relations between identification decisions in both lineups and peak differences in stress indicators. In the Online Supplemental Materials, we provide results of exploratory analyses based on changes in stress responses at four stages of the experimental procedure that corresponded to pre-encoding stress (the math task), encoding stress (the video), post-encoding stress (the presentation), and retrieval stress (the lineup), split for each target lineup character. Findings reported in the Supplemental Materials were consistent with the models reported in the main manuscript.

Table 2 displays proportions of lineup decisions across the experimental conditions and target lineup characters. We also computed compound decision signal detection theory measures of discriminability (d') and bias (c; using the integration rule, see Bruer et al., 2017; Duncan, 2006; Palmer & Brewer, 2012; Palmer et al., 2010).

In a series of generalized linear mixed models (GLMMs, i.e., multilevel logistic regressions) with random intercepts for participant identity and a character-actor identity variable that combined the video character (thief and participant) with actor identity (actors A, B, and C), we first examined any effects of stress (as manipulated or measured) on identification decision accuracy in target-present and target-absent lineups. Next, we looked at stress and accuracy in choosers and non-choosers; and finally, we examined any potential effects of stress on accuracy in low and high similarity lineups. There were three accuracy models (target presence, choosing, and similarity) for each independent variable. We ran separate models for stress condition, BPS, BPD, and STAI (we did not combine multiple stress measures due to their correlation), and a combined model for HR and cortisol (as these two measures did not correlate), resulting in a total of 15 models with 57 p-values. To correct for Type I error, we computed a p-value threshold for a 5% false discovery rate (Benjamini & Hochberg, 1995); therefore, only p-values ≤ .004 were considered significant (see Online Supplemental Materials). We provide a summary of results in the manuscript; full results of all models are reported in Online Supplemental Materials.

TABLE 2 Lineup decisions (proportions), discrimination (*d*'), and response bias (*c*) across stress conditions and actor roles

	Thief lineup							Participant lineup						
			TP			TA				TP			TA	
Group	*d*'	*c*	Hit	FI	Miss	FI	Rej	*d*'	*c*	Hit	FI	Miss	FI	Rej
Low	1.00	−1.00	.36	.45	.18	.69	.31	1.11	−0.69	.35	.38	.27	.55	.45
Mod	0.77	−0.78	.29	.35	.35	.73	.27	0.99	−1.14	.38	.38	.23	.88	.12
High	0.82	−0.66	.28	.44	.28	.60	.40	2.00	−0.80	.55	.10	.35	.56	.44

Abbreviations: FI, filler identification, Rej, lineup rejection; TA, target absent; TP, target present.

4.2.1 | Accuracy and target presence

The models indicated no significant effects of stress condition, measured stress response, or target presence on identification accuracy, *OR*s ≤ 2.16, *zs* ≤ 2.21, *ps* ≥ .027 (see Table SM2 in Online Supplemental Materials).

4.2.2 | Accuracy and choosing

All models indicated that non-choosers made more accurate identification decisions than choosers, *OR*s > 3.60, *zs* > 3.87, *ps* < .001 (see Table SM3 in Online Supplemental Materials). There were no significant effects of manipulated or measured stress, *OR*s ≤ 1.68, *zs* ≤ 2.27, *ps* ≥ .007.

4.2.3 | Accuracy and similarity

The models indicated no significant effects of stress condition, measured stress response, or lineup member similarity on identification accuracy, *OR*s ≤ 1.92, *zs* ≤ 2.19, *ps* ≥ .029 (see Table SM4 in Online Supplemental Materials).

4.2.4 | Confidence-accuracy characteristics

Figure 4 shows the relationship between confidence and accuracy across stress conditions. Accuracy was calculated as the number of guilty suspect identifications divided by the total number of suspect identifications. Total suspect identifications was the number of guilty suspect identifications plus 1/6 of all identifications in target-absent lineups. Only 1/6 of the target-absent identifications were counted as suspects because there were 6 lineup members and in a real case only one of those lineup members would be suspected of the crime. The patterns indicate relatively poor calibration in the high stress group, although our small sample size and associated high variability should be considered when interpreting the data.

4.3 | Choosing

In a series of GLMMs parallel to those described for accuracy, we first examined any effects of stress (as manipulated or measured) on choosing in target-present and target-absent lineups. Then, we examined any potential effects of stress on choosing in low and high similarity lineups. There were two models for choosing (target presence and similarity) for each predictor variable (stress condition, BPS, BPD, and STAI; HR and cortisol were entered together), resulting in a total of 10 models with 38 *p*-values. After a correction for a 5% false discovery rate, only *p*-values < .005 were considered significant.

4.3.1 | Choosing and target presence

The models indicated no significant effects of stress condition, measured stress response, or target presence on choosing, *OR*s ≤ 2.95, *zs* ≤ 2.56, *ps* ≥ .011 (see Table SM5 in Online Supplemental Materials).

4.3.2 | Choosing and similarity

Similarly, there were no significant effect of stress condition, measured stress indicators, or lineup member similarity on choosing, *OR*s ≤ 2.21, *zs* ≤ 2.73, *ps* ≥ .005 (see Table SM6 in Online Supplemental Materials).

4.4 | Statistical power considerations

Before discussing results, we reflect on the possibility that the lack of effects was due to low statistical power. We used package `pwr` (Champely, 2020) to conduct a power analysis for a simplified design assuming two groups (e.g., low and moderate stress groups; to the best of our knowledge, there is no statistical software that could calculate power analysis for proportions in three groups). Assuming a moderate effect of Cohen's $h = 0.50$ with alpha of .004 (consistent with the FDR corrections), our sample when including two lineup responses from each participant had 70% power to detect the effect. If we consider a simplified regression design for the use of measured indicators of stress with a single lineup decision, we likely had over 90% power to detect a moderate effect of $R^2 = 0.15$ (alpha $=$ .004). Therefore, low statistical power could be an issue in the models including stress conditions, but it is a less likely explanation for the null effects in any of the stress indicator models.

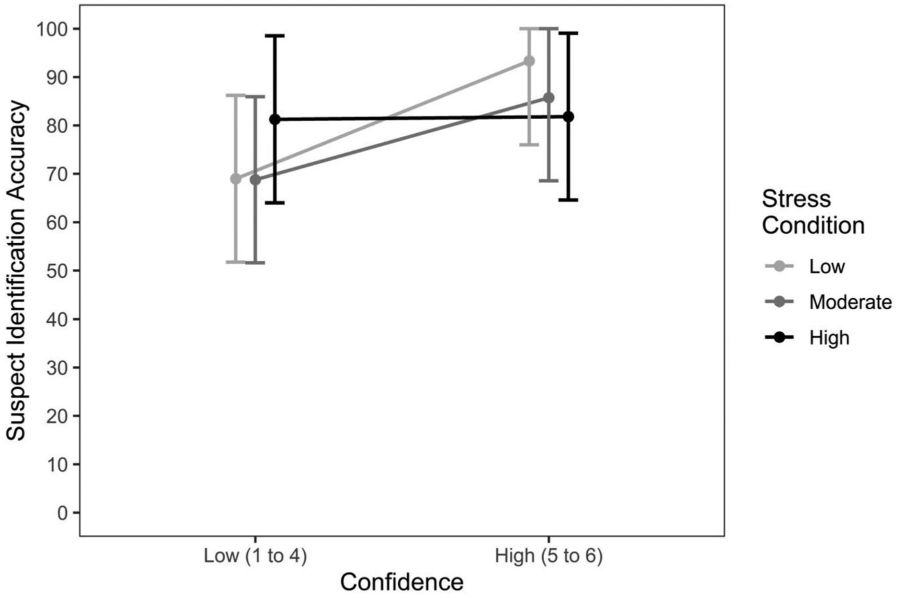

FIGURE 4 Confidence-accuracy characteristic curves across experimental conditions. Confidence levels were collapsed due to small sample sizes. Error bars represent bootstrapped 95% CIs of the means

5 | DISCUSSION

In this study, we investigated the effects of stress on eyewitness identification decisions. We aimed to experimentally induce psychological stress to a moderate and high level, and we measured participants' autonomic and hormonal responses as well as their subjective anxiety. When compared to the low stress group, indicators of stress suggested that participants in the moderate stress group became more stressed in the course of the study, although there were few differences between the moderate and high stress groups. Overall, we found no effects of stress on identification accuracy or choosing.

The absence of a direct effect of stress is consistent with Sauerland et al.'s (2016) finding that, despite the effective induction of stress, identification accuracy was unaffected by whether stress was high or low. However, this null pattern stands in contrast to other prominent studies which indicate a negative effect of stress on identification accuracy (Morgan et al., 2004; Valentine & Mesout, 2009). There are potentially important methodological differences between studies that yielded a negative versus no effect of stress on identification accuracy. For instance, it is likely that the laboratory studies (the current study and Sauerland et al.'s (2016) study) induced a lower stress level, and one that was less immersive and less directly tied to the identification task than the field studies.

A noteworthy contribution of the present data is the variability in the size of correlation between some indicators of stress or subjective anxiety. Across the indicators of autonomic stress response, we found strong correlations between measures of blood pressure, and weak-to-moderate correlations with heart rate. Subjective anxiety demonstrated a moderate-to-large significant relationship with all three autonomic measures: stress increased from baseline in the middle of the procedure and then decreased again, and visual inspection of Figures 2 and 3 reveals that there was greater increase in the moderate and high stress groups than in the low stress group. The peak of the cortisol response showed weak-to-moderate significant correlations with the subjective anxiety and autonomic stress indicators, but there was only a very weak and nonsignificant correlation with heart rate. This high variability in correspondence across various indicators of stress may be helpful in interpreting inconsistent findings across previous studies and should be considered in future designs.

Changes in autonomic response and subjective ratings occur quickly and correspond to a person's current state. In a healthy young person, heart rate recovery can occur within minutes after a stressful experience like the TSST (Kudielka, Buske-Kirschbaum, et al., 2004a; Kudielka, Buske-Kirschbaum, et al., 2004b; Kudielka, Schommer, et al., 2004; see also Figure 2 and Figure SM1 in Online Supplemental Materials). On the other hand, salivary cortisol is estimated to peak approximately 20–30 min after the onset of a stressor (Kudielka, Buske-Kirschbaum, et al., 2004a; Kudielka, Buske-Kirschbaum, et al., 2004b; Kudielka, Schommer, et al., 2004), and considerable time must pass before the effects of cortisol dissipate in the system. Therefore, although autonomic and subjective measures indicated a decline in the stress response, it is reasonable to assume that our participants were still experiencing the impact of the HPA axis stress response during retrieval. In a recent meta-analysis on encoding and retrieval stress in episodic memory, Shields et al. (2017) aimed to disentangle some of the inconsistent effects in the literature and concluded that stress has a negative effect on encoding if (i) the stressor occurs between 11 and 22 min before encoding and (ii) the stressor is not related to the memory task. They also found that stress has a generally negative impact on retrieval. In our study, the onset of stress had a shorter delay to encoding, but the memory task was not related to the stressor, and the stressor also continued after encoding and before retrieval. Despite creating a setting in which memory impairment would be expected, we observed no stress-related effects on identification accuracy or choosing rates.

5.1 | Limitations

Although our participants adhered to the standard restrictions for cortisol testing (e.g., abstaining from caffeine), we included females who were taking hormonal contraception in our sample (see Sauerland et al., 2016, for a similar approach, but see Narvaez Linares et al., 2020 for a thorough discussion on adherence to TSST protocols). Hormonal contraception generally reduces the magnitude of the cortisol response (e.g., Shields et al., 2017). When we excluded participants taking hormonal contraception, we found consistent results in the manipulation check models as well as in the accuracy and choosing models. We should, however, acknowledge the reduction of statistical power in the analysis after the exclusions, so there is still a possibility that an effect could be observed with a larger sample size. It is also the case that, due to methodological requirements of the present study, the 2nd and 3rd salivary cortisol samples were taken approximately 15 and 35 min after the onset of the stress induction, not in the 21–30 min window recommended by Shields (2020). A further limitation is that participants in the current study experienced the stress induction, witnessed the crime video, and completed the lineup all in the same session (Sauerland et al., 2016). Although our autonomic measures indicated that the effects of the stress induction may have dissipated by the stage in the experimental procedure that the lineups were administered, it is possible that participants were still affected by the stress induction at the time of the lineup task. Disentangling effects of stress at encoding and retrieval is critical, particularly given evidence that stress that takes place at these different phases may have opposite effects on memory (Shields et al., 2017). Conducting research with longer retention intervals is required for furthering our understanding of the effect of stress on identification accuracy and to more closely replicating real-world witnessing conditions (see Marr, Sauerland, et al., 2021).

6 | CONCLUSION

We found no effects of stress on the accuracy or the willingness to choose from a lineup. It is possible that our procedure induced relatively low levels of stress, although autonomic, hormonal, and subjective indicators of stress suggested that our stress manipulation was (at least partly) successful. Importantly, there was variability in the correspondence between measured indicators of stress, which should be considered in future designs and interpretation of the extant literature. To disentangle the mixed findings related to the effect of stress on eyewitness memory in the growing literature, researchers should continue to develop novel methodologies to induce stress in the laboratory settings that would more closely approach an eyewitness' experience of stressful events.

ACKNOWLEDGMENTS

This research was funded by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to the first author, Canada Foundation for Innovation grants to the first and second authors, and a University of Regina Dean's Research Award to the first two authors. The authors thank the following research assistants: Denis Gavigan, Dallas Novakowsky, Mark Adkins, Mackenzie Kent, Odell Tan, Jessica Scott, Yuijin Zou, Dominique Dupuis, and Allison Owens.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data are available at: osf.io/37mp2

ORCID

Heather L. Price https://orcid.org/0000-0001-6109-6198
Bianca Hatin https://orcid.org/0000-0002-1329-6010

ENDNOTE

[1] Three exceptions were tested at 10:00 am.

REFERENCES

Alexander, J. K., Hillier, A., Smith, R. M., Tivarus, M. E., & Beversdorf, D. Q. (2007). Beta-adrenergic modulation of cognitive flexibility during stress. *Journal of Cognitive Neuroscience*, *19*(3), 468–478. https://doi.org/10.1162/jocn.2007.19.3.468

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*, 456. https://doi.org/10.3389/fpsyg.2017.00456

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: Series B (Methodological)*, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Birkett, M. A. (2011). The Trier social stress test protocol for inducing psychological distress. *Journal of Visualized Experiments*, *56*, 3238. https://doi.org/10.3791/3238

Britt, T. W. (2005). The effects of identity-relevance and task difficulty on task motivation, stress, and performance. *Motivation and Emotion*, *29*(3), 189–202. https://doi.org/10.1007/s11031-005-9441-3

Bruer, K. C., Fitzgerald, R. J., Price, H. L., & Sauer, J. D. (2017). How sure are you that this is the man you saw? Child witnesses can use confidence judgments to identify a target. *Law and Human Behavior*, *41*, 541–555. https://doi.org/10.1037/lhb0000260

Callister, R., Suwarno, N. O., & Seals, D. R. (1992). Sympathetic activity is influenced by task difficulty and stress perception during mental challenge in humans. *Journal of Physiology*, *454*, 373–387.

Champely, S. (2020). pwr: Basic Functions for Power Analysis. R package version 1.3-0. https://CRAN.R-project.org/package=pwr

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science*, *7*, 238–259. https://doi.org/10.1177/1745691612439584

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior*, *28*, 687–706. https://doi.org/10.1007/s10979-004-0565-x

Duncan, M. J. (2006). *A signal detection model of compound decision tasks*. (Tech. Rep. No. No. TR2006-256). Toronto, ON: Defence Research and Development Canada.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*, 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law*, *19*(2), 151–164. https://doi.org/10.1037/a0030618

Hellhammer, J., & Schubert, M. (2012). The physiological response to Trier social stress test relates to subjective measures of stress during but not before or after the test. *Psychoneuroendocrinology*, *37*, 119–124. https://doi.org/10.1016/j.psyneuen.2011.05.012

Hope, L., Lewinski, W., Dixon, J., Blocksidge, D., & Gabbert, F. (2012). Witnesses in action: The effect of physical exertion on recall and recognition. *Psychological Science*, *23*, 386–390. https://doi.org/10.1177/0956797611431463

Kudielka, B. M., Buske-Kirschbaum, A., Hellhammer, D. H., & Kirschbaum, C. (2004a). Differential heart rate reactivity and recovery after psychosocial stress (TSST) in healthy children, younger adults, and elderly adults. The impact of age and gender. *International Journal of Behavioral Medicine*, *11*, 116–121. https://doi.org/10.1207/s15327558ijbm1102_8

Kudielka, B. M., Schommer, N. C., Hellhammer, D. H., & Kirschbaum, C. (2004). Acute HPA axis responses, heart rate, and mood changes to psychosocial stress (TSST) in humans at different times of day. *Psychoneuroendocrinology*, *29*(8), 983–992. https://doi.org/10.1016/j.psyneuen.2003.08.009

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier social stress test': A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*(1–2), 76–81.

Kudielka, B. M., Buske-Kirschbaum, A., Hellhammer, D. H., & Kirschbaum, C. (2004b). HPA axis responses to laboratory psychosocial stress in healthy elderly adults, younger adults, and children: Impact of age and gender. *Psychoneuroendocrinology*, *29*(1), 83–98. https://doi.org/10.1016/s0306-4530(02)00146-4

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*, 54–64. https://doi.org/10.1038/nrn1825

Marr, C., Otgaar, H., Sauerland, M., Quaedflieg, C. W., & Hope, L. (2021). The effects of stress on eyewitness memory: A survey of memory experts and laypeople. *Memory & Cognition*, *49*, 401–421. https://doi.org/10.3758/s13421-020-01115-4

Marr, C., Sauerland, M., Otgaar, H., Quaedflieg, C. W. E. M., & Hope, L. (2021). The effects of acute stress on eyewitness memory: An integrative review for eyewitness researchers. *Memory*, *29*, 1091–1100. https://doi.org/10.1080/09658211.2021.1955935

Morgan, C. A., Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., & Southwick, S. M. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International Journal of Law and Psychiatry*, *27*, 265–279. https://doi.org/10.1016/j.ijlp.2004.03.004

Narvaez Linares, N. F., Charron, V., Ouimet, A. J., Labelle, P. R., & Plamondon, H. (2020). A systematic review of the Trier social stress test methodology: Issues in promoting study comparison and replicable research. *Neurobiology of Stress*, *13*, 100235. https://doi.org/10.1016/j.ynstr.2020.100235

Oriet, C., & Fitzgerald, R. J. (2018). The single lineup paradigm: A new way to manipulate target presence in eyewitness identification experiments. *Law and Human Behavior*, *42*, 1–12. https://doi.org/10.1037/lhb0000272

Paananen, M., O'Sullivan, P., Straker, L., Beales, D., Coenen, P., Karppinen, J., & Smith, A. (2015). A low cortisol response to stress is associated with musculoskeletal pain combined with increased pain sensitivity in young adults: A longitudinal cohort study. *Arthritis Research & Therapy*, *17*(1), 355. https://doi.org/10.1186/s13075-015-0875-z

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior*, *36*, 247–255. https://doi.org/10.1037/h0093923

Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology: Applied*, *16*, 387–398. https://doi.org/10.1037/a0021034

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rush, E. B., Quas, J. A., Yim, I. S., Nikolayev, M., Clark, S. E., & Larson, R. P. (2014). Stress, interviewer support, and children's eyewitness identification accuracy. *Child Development*, *85*, 1292–1305. https://doi.org/10.1111/cdev.12177

Sauerland, M., Raymaekers, L. H. C., Otgaar, H., Memon, A., Waltjen, T. T., Nivo, M., Slegers, C., Broers, N. J., & Smeets, T. (2016). Stress, stress-induced cortisol responses, and eyewitness identification responses. *Behavioral Sciences & the Law*, *34*, 580–594. https://doi.org/10.1002/bsl.2249

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Shields, G. S. (2020). Stress and cognition: A user's guide to designing and interpreting studies. *Psychoneuroendocrinology*, *112*, 104475. https://doi.org/10.1016/j.psyneuen.2019.104475

Shields, G. S., Sazma, M. A., McCullough, A. M., & Yonelinas, A. P. (2017). The effects of acute stress on episodic memory: A meta-analysis and integrative review. *Psychological Bulletin*, *143*, 636–675. https://doi.org/10.1037/bul0000100

Smeets, T., Cornelisse, S., Quaedflieg, C. W. E. M., Meyer, T., Jelicic, M., & Merckelbach, H. (2012). Introducing the Maastricht acute stress test (MAST): A quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. *Psychoneuroendocrinology*, *37*, 1998–2008.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory*. Consulting Psychologists Press.

Valentine, T., & Mesout, J. (2009). Eyewitness identification under stress in the London dungeon. *Applied Cognitive Psychology*, *23*, 151–161. https://doi.org/10.1002/acp.1463

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.

Wand, G. S., Oswald, L. M., McCaul, M. E., Wong, D. F., Johnson, E., Zhou, Y., Kuwabara, H., & Kumar, A. (2007). Association of amphetamine-induced striatal dopamine release and cortisol responses to psychological stress. *Neuropsychopharmacology*, *32*, 2310–2320. https://doi.org/10.1038/sj.npp.1301373

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Yanagida, T. (2020). misty: Miscellaneous Functions 'T. Yanagida'. R package version 0.3.2. https://CRAN.R-project.org/package=misty

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Price, H. L., Tottenham, L. S., Hatin, B., Fitzgerald, R. J., & Rubínová, E. (2022). Effects of stress on eyewitness identification in the laboratory. *Applied Cognitive Psychology*, *36*(1), 191–202. https://doi.org/10.1002/acp.3910