Being observed caused physiological stress leading to poorer face recognition

Peter J Hills, Daniel Dickinson, Leia M Daniels, Charlotte A Boobyer, & Ross Burton

Bournemouth University

Department of Psychology, Bournemouth University, Talbot Campus, Fern Barrow, Poole, Dorset, BH12 5BB

Abstract

Being observed when completing physical and mental tasks alters how successful people are at completing them. This has been explained in terms of evaluation apprehension, drive theory, and due to the effects of stress caused by being observed. In three experiments, we explore how being observed affects participants' ability to recognise faces as it relates to the aforementioned theories - easier face recognition tasks should be completed with more success under observation relative to harder tasks. In Experiment 1, we found that being observed during the learning phase of an old/new recognition paradigm caused participants to be less accurate at the test phase than not being observed. Being observed at test did not affect accuracy. We replicated these findings in an line-up type task in Experiment 2. Finally, in Experiment 3, we assessed whether these effects were due to the difficulty of the task or due to the physiological stress being observed caused. We found that while observation caused physiological stress, it did not relate to accuracy. Moderately difficult tasks (upright unfamiliar face recognition and inverted familiar face recognition) were detrimentally affected by being observed, whereas easy (upright familiar face recognition) and difficult tasks (inverted unfamiliar face recognition) were unaffected by this manipulation. We explain these results in terms of the direct effects being observed has on task performance for moderately difficult tasks and discuss the implications of these results to cognitive psychological experimentation.

Key words: Face recognition; social pressure; ethnicity; stress; face-inversion effect; ecological validity.

Being observed caused physiological stress leading to poorer face recognition

Face recognition is an expert skill that plays an essential role in eyewitness testimony (Werner, Kühnel, & Markowitsch, 2013). Despite being a form of evidence that is heavily relied-upon in criminal investigations, eyewitness testimony is notably unreliable (Valentine & Davis, 2015). Eyewitness misidentification has been identified as the cause of 29% of the 2068 overturned wrongful convictions documented between 1989 and July 2017 in the US (National Registry of Exonerations, 2017). DNA-based exoneration data reveal that eye-witness misidentification was responsible for conviction in 70% of cases. Both estimator (factors beyond the criminal justice system's control) and system (aspects of the criminal justice system's procedures that affect eye-witness accuracy) variables are known to influence the reliability of eye-witnesses (Mickes & Gronlund, 2017; Wells, 1978). Estimator variables include aspects of the crime (such as the stress and arousal it elicits; Aharonian & Bornstein, 2008; Valentine & Mesout, 2008) and of the witness (including their ability to cope with stress). System variables include instructions given to witnesses during line-up procedures and the behaviour of the police officers during such a procedure (Leach, Cutler, & Van Wallendael, 2009; Mickes & Gronlund, 2017; Valentine & Davis, 2015).

It is common practice for the police to accompany and observe an eye-witness during line-up tasks (either looking through photographs or during the video-line-up procedure). Such observation of people completing face recognition tasks is common place in further applied settings such as at identity checkpoints. The mere presence of an observer can alter both social behaviours and task performance, including memory recall (Berger et al., 1981; Echterhoff & Hirst, 2009). This might be due to the anxiety caused by being observed or due to the increased apprehension of being evaluated during the task. Indeed, during applied eye-witness procedures, witnesses may feel they are expected to produce a (correct) identification which may cause evaluation apprehension (Henchy & Glass, 1968). Given that human behaviour is situationally dependent and as such, the constraints found in highly controlled lab-based experiments risk misinterpreting or overlooking key elements of social-cognitive function (Kingstone, Smilek, & Eastwood (2008) especially in face recognition (Logie, Baddeley, & Woodhead, 1987), it is important to explore how being observed affects face recognition.

Performance can be either enhanced or impaired due to the presence of others. Social facilitation (Allport, 1920; Miyazaki, 2013; Zajonc, 1965) is where enhanced performance is observed in the presence of others. Online gaming research indicates that crowd observation correlates with an increase in player performance (Bowman, Weber, Tamborini & Sherry, 2013). Facilitation extends beyond gaming and into non-gaming cognitive abilities such as hand-eye coordination and mental rotation. However, this effect was only observed in gamers who were considered better than average. Moreover, such outcomes were only present in low-difficulty games: games requiring advanced cognitive ability resulted in a significant increase in failure whilst observers were present indicating social inhibition (when performance is lower due to observation, Pessin, 1933; Hoyt, Blascovich & Smith, 2003).

Zajonc (1965) attempted to explain these two social effects with Drive Theory: the presence of a passive audience leads to greater performance for well-learnt or easy tasks as the individual is utilising their rehearsed behaviours. However, the presence of audiences may inhibit task performance for novel or complex tasks. Thus, the degree of facilitation depends upon a combination of task difficulty and individual skill (Liu, Carmen & Yeung, 2015). Evolutionary benefits appear to form the foundation of this theory; poor awareness of how others will respond to our behaviour makes it advantageous for individuals to be aroused in the presence of others. Our instinctive drive to notice and react with other beings enables greater interaction and social skills.

Alternatively, Cottrell (1968) proposed that an evaluation apprehension (stress concerned with observation and scrutiny of others) may elicit anxiety resulting in performance deterioration. Evaluation apprehension is supported by theories concerning the effects of anxiety and stress upon performance (Eysenck & Calvo, 1992). Stress can present itself in two ways; Eustress and Distress.

When an individual experiences eustress, they have increased feelings of motivation and excitement, resulting in improved performance (Fevre, Matheny, & Kolt, 2003). Distress is experienced when the individual has concern or anxiety about the situation, which is outside of their coping abilities. This results in a decrease in performance and can lead to mental and physical problems (Lazarus & Folkman, 1984). An individual's ability to cope with stressors may result from individual differences (Lazarus & Folkman, 1987). For example, in sport, the presence of another person is deemed as a stressor that can both enhance and reduce individual performance (Manning & Preston, 2003) depending on the level of the stress according to the Yerkes-Dodson Law (Yerkes & Dodson, 1908). The Yerkes-Dodson Law indicates that as stress increases, performance increases to an optimal level before performance negatively correlates with a further rise in anxiety (Martens, & Landers, 1970)

The principle of the Yerkes Dodson Law can also be applied to research investigating stress and face recognition. Yuille, Davies, Gibling, and Marxsen (1994) found that trainee police that were in a higher stress condition recalled more post-event details accurately than those in a lower stress condition in a 1 and 12 week recall test (see also Lindberg, Jones, Collard & Thomas, 2001; Yuille & Cutshall, 1986). Conversely, Valentine and Mesout (2011) found that participants who were more stressed (as measured using physiological measures) were less accurate in a post-event line-up than those who were less stressed. Li, Weerda, Milde, Wolf, and Thiel (2014) found that acute psychological stress affected the recruitment of the medial temporal and frontal areas of the brain leading to impaired face recognition memory (see also Deffenbacher, Bornstein, Penrod & McGorty, 2004, and Avery, Van Der Klok, Heckers, & Blackford, 2016, who found that higher stressed individuals find learning faces more difficult than lower stressed individuals). If we accepted the Yerkes-Dodson Law then we might resolve the apparent contradictions in the above results by suggesting that participants in the Yuille et al. (1994) study were less stressed than those in the Li et al (2014) study. Further, we might apply Drive Theory to these results and suggest that participants who found the face recognition task easier might find that stress enhanced their performance relative to those who found the task more difficult. Indeed, evidence for this latter point stems from Roark, Barrett, Spence, Abdi, & O'Toole (2003) who indicated that stress only affects unfamiliar face recognition. Processing unfamiliar faces is more difficult than processing familiar faces (Hancock, Bruce & Burton, 2000) and may not be based on the same mechanisms (Megreya & Burton, 2005). This would indicate that being observed would enhance face recognition of familiar faces whilst being detrimental to the recognition of unfamiliar faces due to the relative difficulties in processing these stimuli.

A further class of face stimuli that are harder to process are inverted faces. The face-inversion effect is indexed by selectively poorer recognition of inverted faces relative to upright faces as compared to objects (Yin, 1969). It is a highly reliable effect in face recognition (Valentine, 1988) and indicates that upright face recognition is based on expert processing mechanisms not afforded to inverted faces (e.g., Diamond & Carey, 1986). This means that the processing of inverted faces is more difficult than that of processing upright faces (Valentine, 1988) and, given the preceding arguments, might mean that the recognition of inverted faces might be affected differently under observation to the recognition of upright faces.

In this introduction, we have presented evidence to suggest that face recognition performance is likely to be modulated by the social situation under which faces are learnt and subsequently recognised. Observation may lead to social facilitation for easy and well-learnt tasks (Geen, 1983) or may lead to social inhibition for difficult or novel tasks (Amabile, Goldfarb, & Brackfield, 1990). This may be dependent on the level of evaluation apprehension (Cottrell, 1968) or the level of stress that the participant is under (Valentine & Mesout, 2009). Such theorising has important practical implications for the criminal justice system, but also for models of face recognition and cognition more generally. However, empirical evidence for such assertions need to be provided. The series of experiments presented here provide evidence for the effects of observation on face recognition. In Experiment 1, we explore how observation affects face recognition depending on when the observation is at learning or at test. In Experiment 2, we replicate the effects of Experiment 1 but using a line-up procedure. In Experiment 3, we measure whether the effects of observation are due to physiological stress (measured using Galvanic Skin Response and heart rate) or evaluation apprehension (measured using a questionnaire). We utilise different stimuli and different methods in order to demonstrate the generalisability and replicability of these findings.

Experiment 1

In Experiment 1, we employed an old/new recognition paradigm to explore the effects of being observed on face recognition. We devised a scenario where participants believed that they were being observed during the learning phase and/or the test phase (or not at all) of the recognition experiment. Participants believed they were being observed through careful manipulation of the physical environment and instructions given to the participants. In this experiment, we explored the recognition of unfamiliar faces, therefore (due to Drive Theory, Zajonc, 1965, and evaluation apprehension, Cottrell, 1968), we predict that face recognition will be poorer whilst being observed compared to not being observed.

Method

Participants

An opportunity sample of 101 (37 male) Bournemouth University students aged between 18 and 24 years (mean age=21 years) naïve to the purpose of the study were recruited. Participants self-reported that they had normal or correct-to-normal vision and participated as partial fulfilment for a course requirement. Sample size was determined based on an estimate of effect size (r=.46) for how being observed affects false memories (Reysen, 2007) and a Power of .90 (calculated using GPower 3.1).

Materials

Stimuli comprised 38 videos clips (spatial resolution 960x540, frame rate 25fps) of volunteers (19 female, mean age 23.2 years) seated in front of a plain white background in frontal view, with a plain white sheet wrapped over their shoulders thereby occluding clothing such that only the head and neck were visible. All volunteers were White (by self-report), without any distinctive or extraneous features (such as beards or jewellery). The volunteers were recorded counting from 1 to 5. The videos were edited in order to be the same duration (5s), achieved by selecting from the larger database, volunteers who spoke at a similar rate and incorporating a short period of silence at either end of the counting. All videos were recorded using a Panasonic HDC-SD5 high definition video camera under the same ambient illumination and setting. A still image (165x95mm) of each person in a neutral expression with mouth closed was used for the test stimuli.

The experimental procedure was implemented using E-Prime Professional 2 program (Psychology Software Tools Inc., 2010). Stimuli were displayed on a 15' LCD high resolution colour monitor (1280x1024 dpi, refresh rate 50hz). Responses were collected on a standard keyboard. To ensure the manipulations were believable, a HP HD 2300 Webcam and microphone was appropriately positioned in front of the participant with the webcam power light on (switched on, in device settings). A second control computer was positioned facing away from the participant but towards the researcher, this acted as the device the webcam would project onto (with a Skype application open).

Design

A 2x2 design was employed with the factors of observation belief (observed/unobserved) at learning and at test. Both stages of observation were manipulated between subjects. Participants were randomly assigned to one of the four resulting conditions with the proviso that the conditions had a roughly equal number of participants in each. Recognition accuracy was measured using the signal detection theory (Green & Swets, 1966) measure of stimulus discriminability, *d′*. Response times and response bias, *C*, were also recorded. Face stimuli were counterbalanced across participants such that each face appeared as a target and as a distractor an approximately equal number of times.

Procedure

Participants undertook the same experimental procedure (an old/new recognition experiment with three consecutive phases: learning, distractor, and test) with the exception of the instructions provided to the participants. If taking part in an "observed" trial, the participant was informed of the webcam and that it would be recording their behaviour during the trial in order for the Experimenter to establish how people react when viewing faces - participants were not told that their performance was being judged by the camera. A vital element to the success of this study was the belief of observation. In order to make this manipulation believable to participants, the researcher reinforced the realism of the "working" camera by adjusting the zoom, fixing the location and by turning it on in front of the participant. The researcher also showed the participant their face on the secondary computer screen.

Participants in the recorded at learning conditions were informed of the recording at the start of the experiment. Participants in the recorded at learning only condition were told immediately after viewing the learning set of faces that the webcam was no longer recording - the researcher then disconnected the webcam to solidify belief of authenticity. Participants in the recorded at test phase only condition were informed of the recording before the test phase and not before.

During the learning phase, participants viewed nineteen 5 second videos of volunteers counting from 1 to 5 sequentially, in a random order. Participants were required to count aloud from 1 to 5 at the same time as the person in the video. This was to ensure adequate attention was paid to the stimulus. Between each face there was an inter-stimulus interval of 5 seconds; during which the participants were presented with the following participant number and a statement requesting to press any key to continue.

Immediately following this, the distractor phase began, consisting of two 5 minute activities, irrelevant from the task measured. In the first phase, participants viewed the screen populated with multiple (1-15), randomly positioned squares and triangles and were required to indicate using the keyboard (S-square, T-triangle) which was the majority shape. In activity two, simple maths problems were displayed requiring the participant to indicate whether the answer provided was true or false, again by pressing the appropriate key (Z-true, M-false). The order of trials within these tasks was randomised. Immediate feedback was provided for both activities, where a green (for correct) and red (for incorrect) screen with the words “correct” or “incorrect” in the top left corner of the screen was presented; added to this display was also the overall percentage score that the participant had achieved in the task.

Following this, a two-alternative forced-choice recognition test was completed in which all 38 still images of faces were presented in a random order. Nineteen were target images (faces displayed as videos in the learning phase) and 19 were non-targets (new images). Participants were required to indicate whether they had seen the image before or not, using the keypad (S-seen, N-not seen). Test images persisted on screen until a response was made. Between each face there was a plain white screen for 150ms.

Finally, participants were thanked for their involvement in the study and were debriefed fully by the researcher. Ethical approval was granted by Bournemouth University's Faculty of Science and Technology's Research Ethics Committee.

Results

Participants responses were converted into the Signal Detection Theory (Green & Swets, 1966) measures of *d′* (accuracy) and *C* (response bias) using the MacMillan & Creelman (2005) method. A minimum accepted response time of 800ms was set in order to avoid a bias of hyper-clicking during test phase. Analysis revealed four participants who had the majority of their responses below this threshold and were removed from the sample. A further four participants were also removed from the sample due to a large negative *d'* estimate indicating a misunderstanding of the response keys. Excluded participants came from all conditions in a random fashion. Removing these participants did not alter the pattern of the data (decreasing the mean accuracy of the significant effect by 0.08 and 0.04 for the observed at learning and not observed at learning conditions respectively). The subsequent data is summarised in Figure 1 and subjected to a 2 x 2 between-subjects ANOVA.

For recognition accuracy, there was a main effect of being observed at learning, $F(1,89)=4.88$, $MSE=0.42$, $p=.030$, $\eta_p^2=.05$. Recognition accuracy was higher when participants did not believe they were being observed ($M=1.73$, $SE=0.09$) compared to when they believed they were being observed ($M=1.43$, $SE=0.10$). The main effect of being observed at test was not significant, $F(1,89)=0.29$, $MSE=0.42$, $p=.593$, $\eta_p^2<.01$, nor was the interaction, $F(1,89)=0.24$, $MSE=0.42$, $p=.626$, $\eta_p^2<.01$.

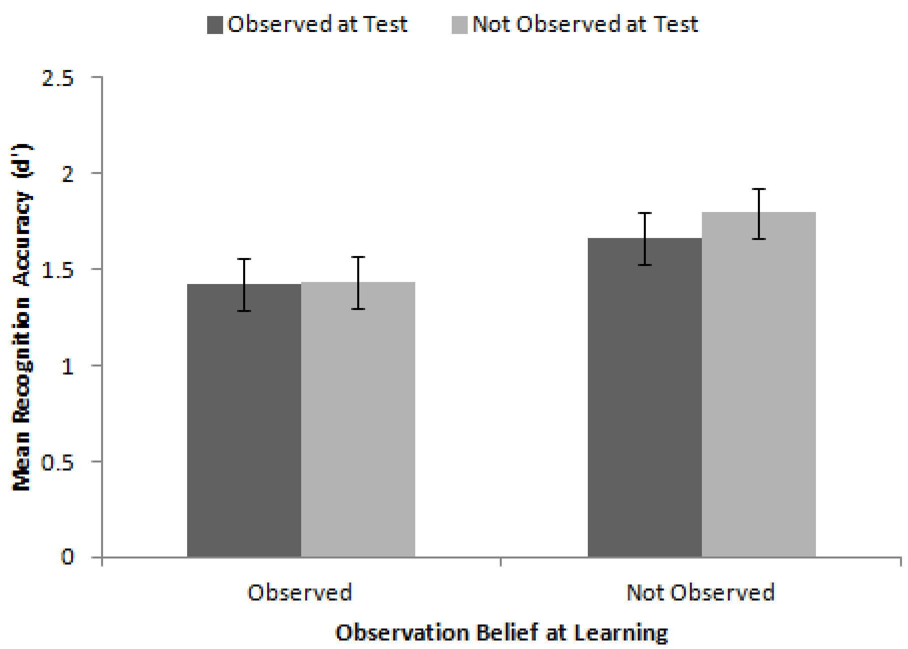

Figure 1. Mean recognition accuracy (*d'*) when participants believed they were being observed at learning and at test. Error bars represent standard error of the mean.

There were no effects of observation belief on response bias, largest F=2.05, smallest p=.156. Nor were there any effects on response time, largest F=1.80, smallest p=.183.

While this approach allows us to explore the locus of when being observed affects face recognition. However, recoding the data may allow us to establish if different amounts of stress (different amounts of observation) differentially affect face recognition performance. Exploring Figure 1 closely, it certainly seems plausible that in the condition where participants were not observed in either condition, recognition performance was higher than when participants were observed in one condition and this was higher than when observed in both conditions. This pattern was directly tested in a subsequent post-hoc one-way ANOVA with the factor coded as fully observed (M=1.39, SE=0.14), partially observed (M=1.56, SE=0.10), and not observed (M=1.80, SE=0.13). This did not meet the traditional level of significance, $F(2,90)=2.34$, MSE=0.42, p=.102, $\eta_p^2=.05$.

Discussion

The results from Experiment 1 highlight that being observed was detrimental to face learning, given that the main effect of observation belief at learning was significant whereas the observation belief at test was not. Being observed potentially caused our participants to find the task more difficult due to the fact that they were under extra pressure caused by the observation. This might have led to more stress and/or more evaluation apprehension leading to poorer performance. The pattern of our secondary analysis revealed that conditions with more stress did lead to lower performance than with less stress, though not significantly. The lack of significance may be related to the fact that the effect of observation was primarily on learning and not at test, or that the incremental amount of stress caused by being observed at test as well as learning is not linear. Nevertheless, the pattern is indicative of increased stress leading to poorer performance. We will return to this issue in Experiment 3.

Experiment 2

Using an old/new recognition paradigm, Experiment 1 has demonstrated that participants being observed during the learning phase show poorer recognition than participants who were not observed during the learning phase. Experiment 2 took this further to investigate the potential effects of observer presence on face identification performance in a line-up procedure. This serves as a replication of the primary finding of Experiment 1 (that being observed influences recognition) whilst also demonstrating whether the effect is generalisable to a more ecologically-valid paradigm. In this Experiment, we did not compare the effect of being observed at learning relative to test as we have established the effect is one of learning. Further, the effects are greater if both phases are observed. Therefore, in future experiments, participants were observed throughout the Experiment. In Experiment 2, participants assumed the role of witnesses to a fictitious, non-violent theft, and attempted to identify the “culprit” of this "crime" from a mock video line-up sequence. We used target-present and target-absent identity parades since both scenarios occur in real-life practice (Steblay, Dysart, & Wells, 2011). We also asked participants to complete questionnaires measuring their self-reported anxiety due to being observed or evaluated. We predict that the presence of an observer during the mock identification procedure would produce increased (self-perceived) social pressure and lower identification accuracy relative to not being observed.

Method

Participants

An opportunity sample of 147 participants (95 female, aged between 18-80 years of age, M = 26.11, SD = 12.77) were recruited. Participants volunteered by responding to adverts placed on social media and around University campus and halls of residence. Participants were compensated with £4 for taking part in this study. Sample size was increased from Experiment 1 because of the change in design, and an estimated reduction in effect size to r = .36. Participants were randomly allocated to one of four conditions based on the line-up type (target-present and target-absent) and whether they would be observed during the task or not.

Materials

The materials used in this research included a crime video, two video line-ups and battery of questionnaires. The crime video and video line-ups were presented to participants using a Tablet PC displayed onto a high-resolution colour LCD monitor (1280x1024 dpi, refresh rate 50hz).

Crime video

The crime video created for this study was filmed in an empty seminar room at Bournemouth University using a stationary digital video camera. This video depicted a non-violent phone theft. The scene began with the "victim" working at a computer, with his bag on the floor next to him and his phone to the left of him on the desk. The "victim" ordered the computer document to print, and went to collect his work from another room – leaving his possessions behind and unattended. After 15 seconds, the "culprit" entered the room, picked up the "victim's" phone and placed it into his pocket, and walked out. The "victim" returned to the room 20 seconds later, noticed his phone was gone, and began to look for it in his bag as the video ended. The "culprit" was visible in the video for a total of 16 seconds.

The video was subsequently edited using Adobe Premiere Pro CC 2015 (Adobe Systems Incorporated) video editing software. The resulting video was 1 minute 28 seconds in length, filmed at a rate of 29 frames per second and was 1920 x 1080 pixels in dimension. Additionally, the audio was removed, the display was cropped to cut out unnecessary background images, and the video quality was degraded slightly to make the "culprit's" features less obvious to the viewer.

Video line-ups

Two video line-up (one target-present and one target-absent) sequences were prepared using the instructions for the National VIPER Bureau video line-up system used by some of the UK Police Forces. Both sequences consisted of nine silent, 15s video clips, each featuring a different individual (the target and 8 fillers or 9 fillers depending on line-up type). The nine filler actors featured in the video line-up sequences were recruited from a local drama department. All were matched using a verbal description to the culprit in accordance with Code D of the Police and Criminal Evidence [PACE] Act (1984; Home Office, 2017).

The 15s line-up clips were recorded in an empty seminar room, with the use of a static digital video camera held at 1.5 m distance from the face. The framing ensured that for each actor the very top of their shoulders was visible and two inches above the actor's head. They were all filmed at a rate of 50 frames per second and a size of 1920 x 1080 pixels. All the actors wore plain black t-shirts to minimise potentially confounding distinguishing features. The timings of all the clips were highly controlled: Each clip began with the actor facing forward and looking directly at the camera for 4s, before turning their head to the left, then to the right (at 7s), finally returning to facing front at 10s.

These clips were subsequently edited using iMovie (Apple Inc.) video editing software. The nine clips were arranged in a random order with a number playing before each clip for 3s. In the final video line-up, the sequence was repeated twice, consistent with PACE. The final versions of video line-up sequences were 6 minutes and 9 seconds long (including instruction screens detailing the format of the sequence and that the "culprit" may or may not be present within it, consistent with PACE).

Questionnaires

Three standardised questionnaires were chosen to record participants self-perceived measures of social pressure: the Liebowitz Social Anxiety Scale (LSAS: Liebowitz, 1987); the Brief Fear of Negative Evaluation-II Questionnaire (BFNE-II: Carleton, Collimore, & Asmundson, 2007); and the Classroom Anxiety Measure (CAM: Richmond, Wrench, & Gorham, 2001) with adapted instructions. These measures are suggested to be both reliable and valid in measuring social anxiety and the effects of being judged (e.g., Carleton et al., 2007; Fresco et al., 2001; Richmond et al., 2001). The CAM was chosen above other methods as it better reflects how participants will react to being assessed rather than a trait anxiety level. The LSAS is a 24-item measure that asks respondents to rate on a scale of 0-3 (3 being the highest) how much fear or anxiety they feel in certain situations (e.g., working under observation). It also asks participants to rate how often they avoid the mentioned situations on a scale of 0-3, with higher scores indicating more avoidance. The BFNE-II is a 12-item measure that explicitly assesses negative evaluation apprehension, in which participants are asked to record how much they agree with each statement (e.g., “I often worry that I will say or do wrong things”) on a scale of 1-5, with higher ratings indicating more agreement. Finally, the CAM is a 20-item state anxiety measure which can be used to record the anxiety participants feel at the moment it is completed. It asks respondents to rate how much they agree with each statement regarding how they are feeling (e.g., “I feel uneasy”) on a scale of 1-5. Like the BFNE-II, higher ratings represent more agreement with the statement. References to the classroom environment were removed from the CAM instructions in order to avoid confusing participants. Given that the responses recorded by the CAM are temporary in nature, this questionnaire was administered to participants twice: first within the battery of questionnaires completed before viewing the allocated video line-up sequence (pre-identification); the second after participants provided their identification decisions (post-identification).

Design

A 2x2 between-subjects design was employed with the variables: observer presence (observer watching or observer absent); and line-up type (target present or target absent). The dependent variable was identification accuracy, confidence, and measures of evaluation anxiety (see materials).

Procedure

After providing informed consent to the research, participants were verbally briefed that they were to assume the role of witnesses to a fictitious phone theft. They watched the crime video once on a laptop either with the experimenter stood next to them watching them or with the experimenter in the corridor outside. This was different to Experiment 1 (in which a camera was used) and was done to better represent how people might observe things (rarely do people get told about being recorded). They were then asked to sit in a nearby waiting area and complete the battery of questionnaires in addition to providing demographic information. Questionnaire order was randomised between participants. On completion, participants were called back into the experiment room.

Participants were verbally briefed regarding the nature of the video line-up in accordance with PACE (Home Office, 2017). They were told that the sequence included nine silent 15-second video clips of different individuals, which would each be identifiable by a number displayed prior to the clip. They were also told that the 9-video sequence will be played twice, and that the “culprit” seen in the crime briefing video may or may not be present in the sequence shown. Participants were instructed to begin playing the sequence by pressing the space bar of the laptop keyboard. Depending on the experimental condition participants were allocated to, the researcher either left the room as the video line-up sequence began playing (for the observer-absent condition) or remained in the room (for the observer-present condition). In instances the researcher was present through the identification procedure, they sat next to participants (as police identification officers do in real-life) and remained neutral in order to avoid posing a distraction or as a source of bias.

On completion of the line-up sequence, participants provided their identification by stating the number of the face or by stating that the culprit was not present. Participants also provided a confidence rating for their identifications on a seven-point Likert scale (with higher ratings indicating higher confidence). Additionally, participants were verbally asked to disclose whether they recognised any of the individuals from the video line-up. No participants knew anyone in the videos. Finally, participants completed the post-identification CAM. Participants were finally thanked and debriefed. Ethical approval was obtained from Bournemouth University's Faculty of Science and Technology's Research Ethics Committee.

Results

As shown in Table 1, participants were significantly more likely to correctly identify the culprit (or identify that a target was not present in target absent line-ups) when they were not being observed than when they were being observed, χ^2=10.10, *p*=.001. Specifically, participants are Ω=3.41 (Odds Ratio) times more likely to make an accurate identification when not being observed than when being observed. Standardised residuals indicate that correct identifications are significantly less likely than expected by chance when being observed. The association between observation and accuracy did not depend on line-up type, χ^2=0.03, *p*=.869. There was also no association between line-up type and accuracy, χ^2=0.22, *p*=.640.

Using independent samples *t*-tests, we compared participants' confidence ratings and measures of their assessment anxiety across observation condition. These revealed that participants were significantly more confident when being observed (*M*=5.07, *SE*=0.15) than when not being observed (*M*=4.59, *SE*=0.15), *t*(145)=2.27, *p*=.025. The post-identification CAM scores indicated higher anxiety when not observed (*M*=49.33, *SE*=1.97) compared to when observed (*M*=42.14, *SE*=1.29), *t*(124.56[1])=3.06, *p*=.003. Indeed, change in CAM scores from pre-identification to post-identification revealed that participants who were not observed (*M*=0.40, *SE*=2.20) showed no change in anxiety compared to a reduction in anxiety in those that were observed (*M*=6.54, *SE*=1.49), *t*(145)=2.32, *p*=.022. No other comparisons were significant, largest mean difference=2.07 , largest *t*(145)=1.19, smallest *p*=.237.

[1] Levene's test of equality of variances was significant, *F*=6.25, *p*=.014, therefore the degrees of freedom were adjusted accordingly.

Table 1.

Count (and expected count in parentheses) for correct and incorrect identifications when participants were observed or not.

		Identification Accuracy	
		Correct	Incorrect
Observation Belief	Observed	12 (20.6)	62 (53.4)
	Not Observed	29 (20.4)	44 (52.6)

Discussion

Participants who were observed were significantly less accurate than participants who were not observed at a line-up task. This replicates and adds to the generalisability of the results found in Experiment 1. Coupled with the lower accuracy, participants who were observed were more confident in their decisions. This further highlights the lack of consistency between confidence and accuracy (Deffenbacher, 1980). This has potentially serious implications for the criminal justice system if witnesses are more confident in their decisions, because they are being observed but their accuracy is actually reduced.

One finding that requires further consideration was that we found participants who were observed were more confident than those who were not. While the Experimenter did not give any facial signals as to whether the participant gave a correct or incorrect response, the participant may have felt they were checking the answers. Given an expectation that people will be corrected if they are providing incorrect information, participants may have felt they were correct because they received no negative feedback. Indeed, participants who seek consensus (even from unhelpful sources) tend to be more confident in their responses than those who do not (Yaniv, Chosehn-Hillel, & Milyavsky, 2009).

In this Experiment, we also assessed participants self-reported response to assessment-related anxiety using three questionnaires. None of these indicated that being observed cause them more stress than not being observed. Instead, post identification, participants who were observed actually reported that they were less stressed than those who were not observed. This indicates a potential release from the stress of observation due to giving a response. Indeed, the change in CAM scores from pre-identification to post-identification show precisely this: participants who were observed showed a reduction in anxiety levels following their identification. It would seem that our observed participants used the identification process to release their anxiety even though their decision was not as accurate as those who were not observed. At first glance, this result might suggest observation alleviates stress in this task. Alternatively, the questionnaires may not accurately pick up subtle physiological changes in stress levels, especially when the source of such stress is not obvious: Participants may have not realised that the observer was increasing their physiological stress. This highlights the dangers of introspection as a basis for performance.

Thus far, we have been advocating a theory that being observed leads to stress that is detrimental to face recognition. The questionnaire results seem to indicate that this thesis may not be entirely valid. Alternatively, it is possible that these data indicate a lack of introspection on the part of our participants. Experiment 3 explores this issue further.

Experiment 3

Experiments 1 and 2 have indicated that being observed during the learning of faces causes subsequent detriment to the recognition of them in either an old/new recognition task or a line-up task. In Experiment 2, we recorded participants' self-reported levels of anxiety and stress using a series of questionnaires. Such questionnaire measurements of stress revolve around the participants self-reporting their feelings about a task. Such reports may be inaccurate due to false memory and ignore physical factors that can be measured during the task (Watson, Pennebaker, & Folger, 1987). Physiological measures, such as GSR and heart rate do not rely on self-report and may better reflect stress. GSR measures the amount of sweat secretion from the sweat glands. This causes changes in the skin conductance. Sweat can be caused by stress (Harker, 2013). Similarly, increased heart rate is indicative of increased stress (Hjortskov, Rissén, Blangsted, Fallentin, Lundberg, & Søgaard, 2004). Both are related to the fight or flight response when confronted with a danger as they prepare the body for increased physical activity (Curtis & O'Keefe, 2002). While it makes sense to be able to remember the danger (if it were a person, for example), when in more extreme danger, it is more sensible to be prepared for a physical response, from an evolutionary perspective. Therefore, one can directly link the stress response to face recognition abilities in addition to considering the effect of stress on cognition more generally.

Therefore, in Experiment 3, we measured participants GSR and heart rate as they completed an old/new recognition paradigm. In order to further assess the predictions of Drive Theory (Zajonc, 1965), we included familiar and unfamiliar upright and inverted faces in this task. Given that upright familiar face recognition is an easy task, we would expect recognition of these faces to be improved due to observation-induced stress. We would expect to see the more difficult tasks to be more affected by the observation-induced stress leading to poorer performance. Similar predictions are made by an evaluation-apprehension account (Cotterill, 1963), therefore we include the evaluation apprehension questionnaire in this study to assess this prediction: Participants scoring high on the evaluation apprehension questionnaire will perform better at the recognition of upright familiar faces and worse on inverted unfamiliar faces. Finally, we have assumed that the link between observation and face recognition is through stress. Given this assumption, we might also expect to see a Yerkes-Dodson pattern in face recognition due to the stress levels measured by GSR and heart rate - i.e., for participants who are less stressed, face recognition generally is less accurate than when participants are optimally stressed. Participants who are very stressed will show poor recognition. We are therefore able to assess three explanations for why observation causes participants to be less accurate at face recognition than those not observed.

Method

Participants

A volunteer sample of 80 (51 female, aged between 18 and 57 years, mean age=27.21 years, SD=9.85) participants took part in this study. Sample size was planned to match that of Experiment 1. Participants responded to adverts placed on social media or participant recruitment websites.

Materials

Two sets of 80 faces were obtained for this study. One set of faces were of familiar people. Two images of UK celebrities were collected from the internet by a research assistant. These were either of singers, actors, TV personalities, and sportsmen and women. These images depicted people in frontal poses in neutral or smiling poses with no jewelry, hats, or other face coverings. The background of the images was masked. These images were presented to 24 undergraduate students from Bournemouth University who did not take part in the current study and were assessed for their familiarity on a seven-point Likert scale and asked to name the face. Only celebrities that were rated above 6 on this scale and were reliably named (or some piece of distinctive biographical information given) by 95% of participants were used in the current study. This led to a final sample of 80 famous faces. These faces were also rated for attractiveness and distinctiveness on seven-point Likert scales by the same participants.

These were matched for number of males, age, attractiveness, and distinctiveness to a set of 80 faces (two versions of each) collected from the Minear and Park (2004) database. These faces were also presented in frontal view, in either a smiling or neutral pose with no jewelry or other distinctive features. Both sets of images were constrained to the size 367 by 425 px in size and presented in 200 dpi resolution. The images were masked such that only the face was visible on a white background (no clothes were visible for either set of images). An inverted set of images was created by rotating the faces by 180^{o} in Microsoft Office Picture Manager.

The experiment was implemented using OpenSesame (version 3.2.4, 2018) presented on a high-resolution colour LCD monitor (1280x1024 dpi, refresh rate 50hz). Participants responded using a standard computer keyboard.

The Evaluation Apprehension Scale (Richmond et al., 2001) was used in this study to measure anxiety. This questionnaire consisted of twenty-two stress-related questions. The participants answered each question on a five-point Likert-type scale, on how much they agreed with the given statement (1 being strongly disagree, 5 being strongly agree). Higher scores indicate more evaluation apprehension. This questionnaire was chosen as it better reflects the construct that we hypothesise explains the effect of observation on face recognition performance than the CAM.

A Nexus4 Galvanic Skin Response (GSR) monitor was used to measure each participants' emotional arousal (GSR). A Polar-watch (POLAR M400) was used to measure participants average heart rate. The participant's heart rate is recorded by the sensor strap and the results fed back to the Polar-watch for the experimenter to see.

Design

A 2x2x2 mixed-subjects design was employed with the between-subjects factor as observation belief (observed or not observed) and the within-subjects factors as face familiarity (famous or unfamiliar) and face orientation (upright or inverted). Physiological stress was used as a covariate. Face recognition accuracy was measured using the Signal Detection Theory (Green & Swets, 1966) measure of stimulus discriminability, *d'*. Response bias was also measured using *C*.

Procedure

After providing informed consent, participants completed seven consecutive phases: set-up, and then two cycles of a learning, distractor, and test phase (one was for famous faces and one was for unfamiliar faces and this was counterbalanced across participants). In the set-up phase, the GSR monitor was attached to the participant's left index finger. The Polar-watch was attached to each participant with the sensor strap placed around the participant's chest, just underneath the chest muscles. Baseline physiological measures were recorded for one minute. Subsequently, participants were told that they were either being recorded via the built-in camera on the monitor (observed condition), or recording was not mentioned at all (not observed condition) in the same way as Experiment 1. Apart from the varied instructions of recording given, all participants followed the same procedure during the experiment. Following the set-up, participants completed the first learning phase.

In the learning phase of the Experiment, participants were informed they were undertaking a face recognition experiment and that they would later be asked to recognise the faces they were being presented with. Participants were presented with 40 (20 inverted) faces sequentially in a random order in the centre of the screen. Participants rated each face for distinctiveness by responding to the question "how easy would this face be to spot in a crowd" on a nine-point Likert-type scale with the anchor points "easy" (distinctive face) and "difficult" (typical face). This was done to ensure attention was paid to the faces (Light, Hollander, & Kayra-Stuart, 1979). Faces were on screen for 2s each. Between each face there was a random-noise mask presented centrally for 250ms.

Following the learning phase, participants either completed the Evaluation Apprehension Scale or a series of demographic questions on the computer. This lasted 60s. Following this, participants completed the recognition test in a similar manner to Experiment 1. All 40 target (20 inverted) and 40 distractor (20 inverted) faces were presented centrally sequentially in a random order. Using "m" for seen before and "z" for not seen before, participants indicated whether the faces was a target or a distractor. Faces were on screen until the participant responded to it. Between each face was a random-noise mask lasting for 250ms.

The sequence of learning, distractor, and test was repeated for the other class of face. On completion, participants were thanked and debriefed. The experiment lasted a total of 20 minutes for each participant to complete. Ethical approval was obtained from Bournemouth University's Faculty of Science and Technology's Research Ethics Committee.

Results

Participant responses were treated in the same way as in Experiment 1 and summary data is shown in Figure 2. The recognition accuracy data were subjected to a 2 x 2 x 2 mixed-subjects ANOVA. This revealed that familiar faces (M=1.52, SE=0.05) were recognised more accurately than unfamiliar faces (M=1.05, SE=0.07), $F(1,78)=39.97$, $MSE=0.58$, $p<.001$, $\eta_p^2=.28$. Upright faces (M=1.45, SE=0.06) were recognised more accurately than inverted faces (M=1.13, SE=0.06), $F(1,78)=26.19$, $MSE=0.32$, $p<.001$, $\eta_p^2=.25$. These factors did not interact, $F(1,78)=0.64$, $MSE=0.32$, $p=.427$, $\eta_p^2=.01$. Observed participants (M=1.15,SE=0.07) were less accurate than participants who were not observed (M=1.42,SE=0.07), $F(1,78)=8.20$, $MSE=0.69$, $p=.005$, $\eta_p^2=.10$.

Critically, the three-way interaction was significant, $F(1,78)=6.98$, $MSE=0.52$, $p=.010$, $\eta_p^2=.08$. A series of t-tests comparing the effect of observation for each condition were run. This was significant for upright unfamiliar faces, $t(78)=2.70$, $p=.008$ and for familiar inverted faces, $t(78)=2.45$, $p=.016$ but not significant for familiar upright faces, $t(78)=0.73$, $p=.470$, nor for unfamiliar inverted faces, $t(78)=0.05$, $p=.964$.

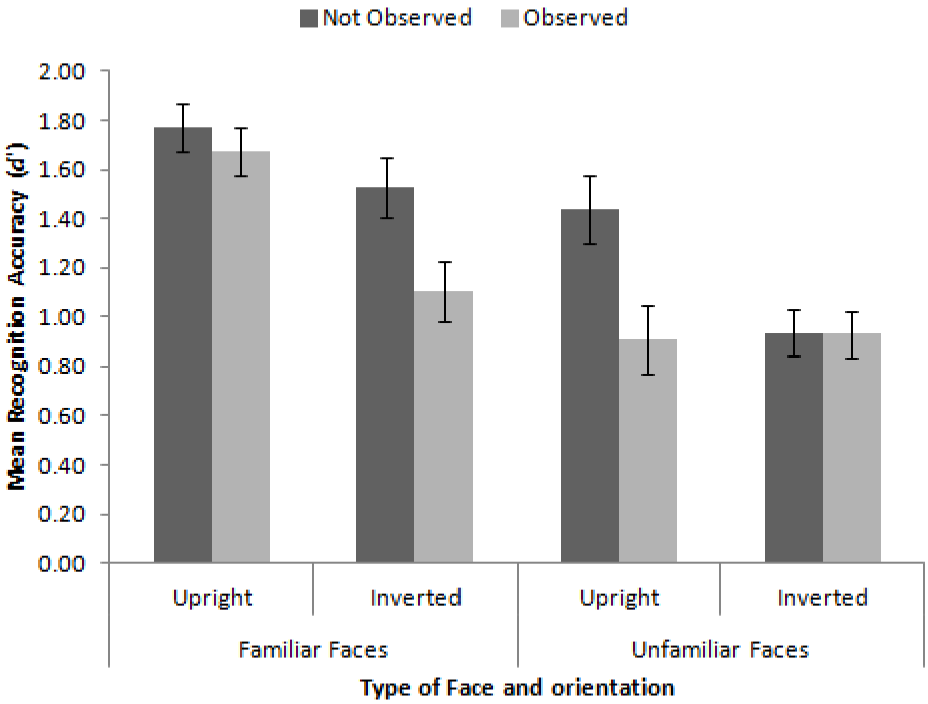

Figure 2. Mean recognition accuracy of familiar and unfamiliar upright and inverted faces for participants who believed they were being observed or not. Error bars represent standard error of the mean.

Our next analysis focused on the measures of stress. We did not find that self-reported evaluation apprehension differed depending on whether participants were observed (*M*=2.82, *SE*=0.10) or not (*M*=2.82, *SE*=0.10), *t*(78)=0.05, *p*=.959. As predicted, average corrected for baseline GSR and heart rate were higher when observed (*M*=7.65, *SE*=0.05,for GSR, and *M*=82, *SE*=1.56, for heart rate) than when not observed (*M*=7.45, *SE*=0.06,for GSR, and *M*=78, *SE*=1.44, for heart rate), *t*(78)=2.49, *p*=.008, one-tailed, for GSR, and *t*(78)=1.93, *p*=.029, one-tailed, for heart rate. We, therefore, assessed whether the observed effects on recognition accuracy due to observation were mediated by physiological stress by running a series of between-subjects ANCOVAs on each type of face with heart rate and GSR recorded at learning and at test entered as separate covariates. Table 2 reports any changes to significance to the accuracy results presented above. There were no changes to the pattern of significance when including these as covariates: Significant effects remained significant and non-significant effects remained non-significant.

A parallel analysis on response bias revealed that there were no effects nor interactions involving the observation factor, largest *F*=2.94, smallest *p*=.090. Nevertheless, participants were more likely to respond with an "old" response to familiar faces (*M*=-.02, *SE*=.03) than unfamiliar faces (*M*=.32, *SE*=.04), *F*(1,78)=66.60, *MSE*=0.14, *p*<.001, η_p^2=.46. Participants were more likely to respond with a "new" response to inverted faces (*M*=.20, *SE*=.03) than upright faces (*M*=.10, *SE*=.03), *F*(1,78)=7.03, *MSE*=0.10, *p*=.010, η_p^2=.08.

Table 2.

Significance level changes for the effect of observation on face recognition accuracy with GSR and heart rate at learning and at test as a covariate.

			Significance with Covariate			
			GSR		Heart Rate	
Face type and orientation		Original Significance value	At learning	At test	At learning	At test
Familiar Faces	Upright	*p*=.470	*p*=.563	*p*=.595	*p*=.576	*p*=.564
	Inverted	*p*=.016	*p*=.021	*p*=.016	*p*=.042	*p*=.042
Unfamiliar Faces	Upright	*p*=.008	*p*=.011	*p*=.035	*p*=.009	*p*=.014
	Inverted	*p*=.964	*p*=.999	*p*=.972	*p*=.993	*p*=.972

The parallel analysis on learning response times (to make distinctiveness judgements) revealed no effects nor interactions involving the observation factor, largest *F*=1.22, smallest *p*=.272. However, unfamiliar (M=1503, SE=122) and upright faces (M=1521, SE=126) were rated faster than familiar (M=1913, SE=167) and inverted faces (M=1895, SE=159), $F(1,78)=21.26$, MSE=631579, $p<.001$, $\eta_p^2=.21$, and $F(1,78)=31.81$, MSE=352145, $p<.001$, $\eta_p^2=.29$, respectively. During recognition, unfamiliar faces (M=996, SE=67) were responded to faster than familiar faces (M=1275, SE=92), $F(1,78)=26.58$, MSE=233552, $p<.001$, $\eta_p^2=.25$. No other effects were significant, largest F=1.77, smallest p=.188.

Discussion

In this study, we attempted to establish whether the effect of observation on unfamiliar face recognition was different to familiar face recognition. We hypothesised that because familiar face recognition is an easier task than unfamiliar face recognition, it is less likely to be affected by stress according to the Yerkes-Dodson Law. Based on the same law, recognition of inverted unfamiliar faces will be less affected by stress. This is what we found in the present study. Indeed, the two conditions that had similar performance levels - inverted familiar and upright unfamiliar faces - were equally affected by the observation manipulation. This would suggest that these tasks are sufficiently difficult to be affected by stress but not too difficult.

While this explanation seems to fit the data well, and we observed physiological effects (increased GSR and heart rate) due to being observed, there was no relation between physiological response and face recognition performance. In other words, while observation caused an increase in physiological stress this did not relate to the detriments observed in face recognition. We also did not observe any effects of being observed on self-reported evaluation apprehension. These results suggest one of two possible alternatives: the physiological measures were not sensitive enough to detect differential stress effects in what is a fairly simple face recognition task; or that the effect observation has on face recognition is not through physiological stress. Nevertheless, we found that face recognition was affected by being observed and that easy and very difficult tasks were less affected by being observed than those that were moderately difficult. This pattern is not the result of floor or ceiling effects and all data were significantly different to chance and ceiling performance. Rather this difference reflects a real psychological effect.

General Discussion

In the current set of experiments, we investigated whether being observed would affect face recognition performance based on the notion that being observed caused changes in several areas of cognition and behaviour (Bowman, Weber, Tamborini, & Sherry, 2013; Berger et al., 1981; Hameroff, Nip, Porter, Tuszynski, 2002; Hoyt et al., 2003; Laidlaw Foulsham, Kuhn, & Kignstone, 2011). Across three studies, employing different procedures to enhance the generalisability of these findings, we found that being observed during the initial encoding of faces causes them to be less well recognised subsequently. This effect replicated into a line-up task. This effect was observed when completing moderately difficult face recognition tasks (such as the recognition of upright unfamiliar faces and inverted familiar faces) but not during easier tasks (recognition of familiar faces) or difficult tasks (recognition of inverted unfamiliar faces).

We found that the effects of observation affected accuracy but not response bias nor response time. This suggests that the effect of observation is directly affecting the recognition response but not the threshold applied. Indeed, similar results have been observed by Avery et al (2016). Active ratings of faces and recall from memory may incur a set delay and any effect of observation is negligible on processing speed.

The effect we observed here further reinforces the role of social inhibition upon task performance (Allport, 1920) and appear consistent with the Evaluation Apprehension Model of social inhibition (Cottrell, 1968) and Drive Theory (Zajonc, 1965). Participants were less likely to encode faces as well when being observed, which suggests individuals were experiencing high anxiety at the initial stage of the experiment. This induced anxiety may have somewhat influenced individuals’ ability to learn faces. Learning upright unfamiliar faces is a sufficiently difficult task to be affected by social inhibition. However, learning familiar faces is not a difficult enough task to be affected in this way. One result that was not consistent with Drive Theory is that the most difficult stimuli to process were not affected by observation. While there were no floor effects, we interpret this as the task was too difficult to be further affected by observation. Alternatively, this is due to something more unique to the processing of faces - upright faces are processed in an expert manner, whereas inverted faces are not (Yovel & Kanwisher, 2009). The familiar face recognition manipulation we employed seems to rule out this later explanation, however.

An alternative suggestion for the present data stems from Gobel, Kim, and Richardson (2015) who reported that participants physically altered how they viewed faces once told that they would meet the actors on screen. Potentially, the fear of scrutiny caused participants to adjust gaze, resulting in poorer overall face learning. When eye movements are restricted whilst viewing faces (Henderson, Williams, & Falk, 2005) or are cued to unusual locations of images (Hills, Cooper, & Pake, 2013), face recognition performance is lower than unrestricted viewing. Therefore, it seems plausible that the stress caused by being observed might be altering the encoding of the faces and this is damaging to face recognition.

While we have presented an argument based on anxiety induced through observation causing stress that affects face recognition, our data do not rule out other explanations. Indeed, if the link from observation to accuracy was mediated by stress, we would have assumed that the there would be no significant effect of observation when physiological stress was included as a covariate (Experiment 3). We observed a very small reduction in significance levels, but not a removal suggesting only a partial mediation at best. It is possible that the effect of being observed causes an added distraction to participants during the task. Paying inadequate attention when learning stimuli due to higher perceptual load caused by distraction impairs subsequent recognition for geometric shapes (e.g., Rock, Schauer, & Halper, 1976), line drawings (Goldstein & Fink, 1981; Rock & Gutman, 1981), and words (e.g., Gardiner & Parkin, 1990). Indeed, Baron (1986) highlighted that the attentional system can be overloaded due to distraction caused by the mere presence of other people. The mere presence of an Experimenter impairs attentional maintenance in working memory (Belletier & Camos, 2018), the identification of central cues (Wühr & Huestegge, 2010), and reading span tasks (Belletier, Davranche, Tellier, Dumas, Vidal, Hasbroucq, & Huguet, 2015). Belletier et al. (2015) hypothesise that the presence of an experimenter causes participants to attend to them, limiting their attentional resources. This might cause attentional (response) conflict, leading to poor performance on difficult or attention-demanding tasks (Huguet, Dumas, & Monteil, 2004; Huguet, Galvaing, Monteil, & Dumas, 1999; Muller & Butera, 2007; Normand, Bouquet, & Croizet, 2014; Sharma, Booth, Brown, & Huguet, 2010).

The notion that the mere presence of others might cause an increase in attentional demands, evidence suggests that face recognition performance remains relatively stable under increasing working memory loads (caused by distractor items, Lavie, Ro, & Russell, 2003; Stevenage, Neil, Barlow, Dyson, Eaton-Brown, & Parsons, 2013). However, face recognition can be impaired when another task has to be performed or attention is disrupted (Kellogg, 1980; Kellogg, Cocklin, & Bourne, 1982; Reinitz, Morrissey, & Demb, 1994), consistent with load theory (Lavie, 1995, 2001). It could, therefore, be that the mere fact of being observed changes the context of the Experiment to be a multi-task Experiment, thereby causing a distraction. Limited attentional capacity at learning reduces the number of faces that can later be recognised (Jenkins, Lavie, & Driver, 2005), consistent with this theory. To explain the pattern of results observed in Experiment 3 (whereby familiar faces were not affected by this multi-task context), we can turn to Event-Related Potential (ERP) evidence of Neumann, Mohamed, and Schweinberger (2011). They reported that the recognition of unfamiliar faces was impaired by higher perceptual load, however, familiar faces still produced the N250r ERP, which reflects repetition and familiarity. This suggested that familiar faces could still be processed under higher load. This theoretical position is compatible with Drive Theory (i.e., observation may cause evaluation apprehension and this causes attentional capacity limitations), but offers another level of interpretation.

The preceding commentary is consistent with the notion that stress impacts on brain functioning related to working memory and attention. Stress reduces the recruitment of the brain areas responsible for working memory and attention (the frontal and medial temporal regions; Li et al., 2014), which impacts on the ability to recognise faces. The reason why this effect is primarily on upright unfamiliar faces might relate to findings of Civile, McLaren, and McLaren (2018) who have demonstrated anodal transcranial direct current stimulation to a frontal site (Fp3) reduces the recognition of upright faces. Potentially, stress and anodal stimulation are having similar effects on the frontal regions, thereby reducing upright face recognition.

In the current study, we used two methods to observe participants: through a webcam and by direct observation by close proximity. Both caused comparable effects suggesting that the way the participant appraised the observation was having an effect on their performance. There are known individual difference and personality variables that might affect how participants appraise such observation (Avery et al., 2016). Future studies could explore these to establish if the effect of observation on face recognition is a general one or if some participants are more susceptible or not. Furthermore, while the observation of the participants was direct and obvious and in relatively close proximity, we cannot determine the limits of this observation effect. In gaming research the distance from the player to the observer does have an effect (Hoyt et al., 2003). Here, the observer stated that they were observing the participant or not in addition to observing from a close distance or moving into another room. Therefore, it will be important for future research in cognitive psychology to explore whether passive or direct observation has differential effects on task performance.

This leads to a critical implication of the present work on future studies in face recognition, perception, and cognitive psychology more generally. In many studies, the participants are observed as they complete the task. Potentially, this has modifying effects on task performance. Such effects may be subtle and there may be individual differences in who is affected and to what degree.

Nevertheless, failures to replicate and inconsistent findings in previous literature across psychological research might be, in part, affected by the presence or absence of the experimenter.

There are potential practical implications of the current work, especially for those involved in learning faces or trying to detect faces in the criminal justice system. Anyone learning faces is best advised not to do it whilst under observation or social pressure. Of course, the present studies were laboratory based and while we attempted (in Experiment 2) to match the behaviour of the experimenter to current police practice, there were no negative consequences for the participant as there might be in real eye-witness situations. This may further alter the effect of being observed on face recognition performance.

In conclusion, we have found that face recognition is negatively affected by being observed. This is especially evident during the learning of upright unfamiliar faces (or inverted familiar faces). We hypothesise that this is due to social observation causing participants to be somewhat anxious and due to the Yerkes-Dodson law, performing poorly in a relatively difficult task. However, we did not find that physiological stress mediated the relationship between observation and face recognition. Therefore, we assume that observation has direct affects on how people encode faces and this has a detrimental effect on the recognition of face recognition tasks that are moderately difficult.

References

Aharonian, A. A., & Bornstein, B. H. (2008). Stress and Eyewitness Memory. In Cutler, B. L. (Eds.), *Encyclopedia of Psychology and Law*. Newbury Park, CA: SAGE Publications.

Allport, F. H. (1920). The influence of the group upon association and thought. *Journal of experimental psychology*, 3(3), 159.

Amabile, T. M., Goldfarb, P., & Brackfleld, S. C. (1990). Social influences on creativity: Evaluation, coaction, and surveillance. *Creativity research journal*, *3*(1), 6-21.

Avery, S. N., Van Der Klok, R. M., Heckers, S., & Blackford, J. U. (2016). Impaired face recognition is associated with social inhibition. *Psychiatry research*, 236, 53-57.

Baron, R. S. (1986). Distraction-conflict theory: Progress and problems. In *Advances in experimental social psychology* (Vol. 19, pp. 1-40). Academic Press.

Belletier, C., & Camos, V. (2018). Does the experimenter presence affect working memory?. *Annals of the New York Academy of Sciences*.

Belletier, C., Davranche, K., Tellier, I. S., Dumas, F., Vidal, F., Hasbroucq, T., & Huguet, P. (2015). Choking under monitoring pressure: being watched by the experimenter reduces executive attention. *Psychonomic bulletin & review*, *22*(5), 1410-1416.

Berger, S. M., Hampton, K. L., Carli, L. L., Grandmaison, P. S., Sadow, J. S., Donath, C. H., & Herschlag, L. R. (1981). Audience-induced inhibition of overt practice during learning. *Journal of Personality and Social Psychology*, 40(3), 479.

Bowman, N. D., Weber, R., Tamborini, R., & Sherry, J. (2013). Facilitating game play: How others affect performance at and enjoyment of video games. *Media Psychology*, 16(1), 39-64.

Bowman, N. D., Weber, R., Tamborini, R., & Sherry, J. (2013). Facilitating game play: How others affect performance at and enjoyment of video games. *Media Psychology*, 16(1), 39-64.

Carleton, R. N., Collimore, K. C., & Asmundson, G. J. (2007). Social anxiety and fear of negative evaluation: Construct validity of the BFNE-II. *Journal of Anxiety Disorders, 21*(1), 131-141. DOI: https://dx.doi/10.1016/j.janxdis.2006.03.010

Civile, C., McLaren, R., & McLaren, I. P. L. (2018). How we can change your mind: Anodal tDCS to Fp3 alters human stimulus representation and learning. *Neuropsychologia*, *119*, 241-246.

Cottrell, N. B. (1968). Performance in the presence of other human beings: Mere presence, audience, and affliation effects. In *F. E. Simmel, R. A. Hoppe, & G. A. Milton (Eds.)*, Social Facilitation and Imitative Behaviour. Boston: Allyn & Bacon.

Cottrell, N. B., Wack, D. L., Sekerak, G. J., & Rittle, R. H. (1968). Social facilitation of dominant responses by the presence of an audience and the mere presence of others*. Journal of personality and social psychology*, 9(3), 245.

Curtis, B. M., & O'Keefe Jr, J. H. (2002). Autonomic tone as a cardiovascular risk factor: the dangers of chronic fight or flight. In *Mayo Clinic Proceedings* (Vol. 77, No. 1, pp. 45-54). Elsevier.

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship?. *Law and Human Behavior*, *4*(4), 243.

Deffenbacher, K., Bornstein, B., Penrod, S., & McGorty, E. (2004). A Meta-Analytic Review of the Effects of High Stress on Eyewitness Memory. *Law And Human Behaviour, 28*(6), 687-706. Doi: 10.1007/s10979-004-0565-x

Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General*, *115*(2), 107.

Echterhoff, G., & Hirst, W. (2009). Social influence on memory. *Social psychology*, *40*(3), 106-110.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion, 6*(6), 409-434. Doi: 10.1080/02699939208409696

Fresco, D. M., Coles, M. E., Heimberg, R. G., Leibowitz, M. R., Hami, S., Stein, M. B., & Goetz, D. (2001). The Liebowitz Social Anxiety Scale: A comparison of the psychometric properties of self-report and clinician-administered formats. *Psychological Medicine, 31*(6), 1025-1035. DOI: https://dx.doi/10.1017/S0033291701004056

Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition*, *18*(6), 579-583.

Garrett, B. (2011). *Convicting the innocent: Where criminal prosecutions go wrong.* Cambridge, MA: Harvard University Press. doi:10.4159/harvard.9780674060982

Geen, R. G. (1983). Evaluation apprehension and the social facilitation/inhibition of learning. *Motivation and Emotion*, *7*(2), 203-212.

Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, 136, 359-364.

Goldstein, E. B., & Fink, S. I. (1981). Selective attention in vision: Recognition memory for superimposed line drawings. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(5), 954.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hameroff, S., Nip, A., Porter, M., & Tuszynski, J. (2002). Conduction pathways in microtubules, biological quantum computation, and consciousness. *Biosystems*, 64(1), 149-168.

Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in cognitive sciences*, *4*(9), 330-337.

Harker, M. (2013). Psychological sweating: a systematic review focused on aetiology and cutaneous response. *Skin pharmacology and physiology*, *26*(2), 92-100.

Henchy, T., & Glass, D. C. (1968). Evaluation apprehension and the social facilitation of dominant and subordinate responses. *Journal of personality and social psychology*, *10*(4), 446.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & cognition*, *33*(1), 98-106.

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013). First fixations in face processing: The more diagnostic they are the smaller the face-inversion effect. *Acta Psychologica*, *142*(2), 211-219.

Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European journal of applied physiology*, *92*(1-2), 84-89.

Home Office. (2017, February). Police and Criminal Evidence Act (PACE) – Code D. Revised Code of Practice for the Identification of Persons by Police Officers (PDF Document). Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/592562/pace-code-d-2017.pdf

Hoyt, C. L., Blascovich, J., & Swinth, K. R. (2003). Social inhibition in immersive virtual environments. Presence: *Teleoperators and Virtual Environments*, 12(2), 183-195.

Huguet, P., Dumas, F., & Monteil, J.-M. (2004). Competing for a desired reward in the Stroop task: When attentional control is unconscious but effective versus conscious but ineffective. *Canadian Journal of Experimental Psychology, 58,* 153–167.

Huguet, P., Galvaing, M. P., Monteil, J. M., & Dumas, F. (1999). Social presence effects in the Stroop task: Further evidence for an attentional view of social facilitation. *Journal of Personality and Social Psychology, 77,* 1011–1025.

Jenkins, R., Lavie, N., & Driver, J. (2005). Recognition memory for distractor faces depends on attentional load at exposure. *Psychonomic Bulletin & Review*, *12*(2), 314-320.

Kellogg, R. T. (1980). Is conscious attention necessary for long-term storage?. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(4), 379.

Kellogg, R. T., Cocklin, T., & Bourne Jr, L. E. (1982). Conscious attentional demands of encoding and retrieval from long-term memory. *The American journal of psychology*, 183-198.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, 99(3), 317-340.

Laidlaw, K. E., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences*, 108(14), 5548-5553.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 451.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in cognitive sciences*, *9*(2), 75-82.

Lavie, N., Ro, T., & Russell, C. (2003). The role of perceptual load in processing distractor faces. *Psychological science*, *14*(5), 510-515.

Lazarus, R. S., & Folkman, S. (1984). Coping and adaptation. *The handbook of behavioral medicine*, *282325*.

Lazarus, R. S., & Folkman, S. (1987). Transactional theory and research on emotions and coping. *European Journal of personality*, *1*(3), 141-169.

Leach, A. M., Cutler, B. L., & Van Wallendael, L. (2009). Lineups and eyewitness identification. *Annual Review of Law and Social Science*, *5*, 157-178. DOI: 10.1146/annurev.lawsocsci.093008.131529

Le Fevre, M., Matheny, J., & Kolt, G. S. (2003). Eustress, distress, and interpretation in occupational stress. *Journal of managerial psychology*, *18*(7), 726-744.

Li, S., Weerda, R., Milde, C., Wolf, O. T., & Thiel, C. M. (2014). Effects of acute psychosocial stress on neural activity to emotional and neutral faces in a face recognition memory paradigm. *Brain imaging and behavior*, *8*(4), 598-610.

Liebowitz, M. R. (1987). Social phobia. Modern Problems of Pharmacopsychiatry, 22, 141–173. Questionnaire retrieved from: http://healthnet.umassmed.edu/mhealth/LiebowitzSocialAnxietyScale.pdf on 21/02/2017.

Lindberg, M., Jones, S., Collard, L., & Thomas, S. (2001). Similarities and Differences in Eyewitness Testimonies Of Children Who Directly Versus Vicariously Experience Stress. *The Journal Of Genetic Psychology, 162*(3), 314-333. Doi: 10.1080/00221320109597486

Liu, E. S., Carmen, J. Y., & Yeung, D. Y. (2015). Effects of approach to learning and self-perceived overall competence on academic performance of university students. *Learning and Individual Differences*, 39, 199-204.

Logie, R. H., Baddeley, A. D., & Woodhead, M. M. (1987). Face recognition, pose and ecological validity. *Applied Cognitive Psychology, 1*(1), 53-69.

Macmillan, N. A., & Creelman, C. D. (2010). *Detection theory: A user's guide*. Cambridge, UK: Cambridge University Press.

Manning, D., & Preston, A. (2003). Organizational stress: Focusing on ways to minimize Distress. *Cupa HR Journal*, *54*, 15-18.

Martens, R., & Landers, D. M. (1970). Motor performance under stress: A test of the inverted-U hypothesis. *Journal of Personality and Social Psychology*, 16(1), 29.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & cognition*, *34*(4), 865-876.

Mickes, L., & Gronlund, S. (2017). Eyewitness Identification. In J. H. Byrne (Ed.), *Cognitive Psychology of Memory, Vol 2 (Learning and Memory: A Comprehensive Reference* [2nd Ed]). Oxford: Elsevier.

Miyazaki, Y. (2013). Increasing Visual Search Accuracy by Being Watched. *Plos ONE, 8*(1), e53500. Doi: 10.1371/journal.pone.0053500

Muller, D., & Butera, F. (2007). The focusing effect of self-evaluation threat in coaction and social comparison. *Journal of Personality and Social Psychology, 93,* 194–211.

National Registry of Exonerations. (2017, July). % EXONERATIONS BY CONTRIBUTING FACTOR. Retrieved from: http://www.law.umich.edu/special/exoneration/Pages/ExonerationsContribFactorsByCrime.aspx on 24/07/17

Neumann, M. F., Mohamed, T. N., & Schweinberger, S. R. (2011). Face and object encoding under perceptual load: ERP evidence. *NeuroImage*, *54*(4), 3021-3027.

Normand, A., Bouquet, C. A., & Croizet, J.-C. (2014). Does evaluative pressure make you less or more distractible? Role of top-down attentional control over response selection. *Journal of Experimental Psychology: General, 143,* 1097–1111. doi: 10.1037/a0034985

Pessin, J. (1933). The comparative effects of social and mechanical stimulation on memorizing. *The American Journal of Psychology*, 45(2), 263-270.

Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 161.

Reysen, M. B. (2007). The effects of social pressure on false memories. *Memory & Cognition*, *35*(1), 59-65.

Richmond, V. P., Wrench, J. S., & Gorham, J. (2001). *Communication, Affect, and Learning in the Classroom*. Acton, MA: Tapestry Press. PDF version of book retrieved from: http://jasonswrench.com/pdf/affect_book.pdf

Roark, D. A., Barrett, S. E., Spence, M. J., Abdi, H., & O'Toole, A. J. (2003). Psychological and neural perspectives on the role of motion in face recognition. *Behavioral and cognitive neuroscience reviews*, *2*(1), 15-46..

Rock, I., & Gutman, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(2), 275.

Rock, I., Schauer, R., & Halper, F. (1976). Form perception without attention. *The Quarterly journal of experimental psychology*, *28*(3), 429-440.

Sharma, D., Booth, R., Brown, R., & Huguet, P. (2010). Exploring the temporal dynamics of social facilitation in the Stroop task. *Psychonomic Bulletin & Review, 17,* 52–58. doi: 10.3758/PBR.17.1.52

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-Two Tests of the Sequential Lineup Superiority Effect: A Meta-Analysis and Policy Discussion. *Psychology, Public Policy, and Law*, *17*(1), 99-139. DOI: 10.1037/a0021650

Stevenage, S. V., Neil, G. J., Barlow, J., Dyson, A., Eaton-Brown, C., & Parsons, B. (2013). The effect of distraction on face and voice recognition. *Psychological research*, *77*(2), 167-175.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British journal of psychology*, *79*(4), 471-491.

Valentine, T., & Davis, J. P. (2015). Identification and Surveillance of Facial Images: Progress and Problems. In Valentine, T., & Davis, J. P. (Eds*.), Forensic Facial Identification: Theory and Practice of Identification from Eyewitnesses, Composites and CCTV (*pp. 3-14). Wiley-Blackwell.

Valentine, T., & Mesout, J. (2008). Eyewitness identification under stress in the London Dungeon. *Applied Cognitive Psychology*, *23*(2), 151-161. DOI: 10.1002/acp.1463

Watson, D., Pennebaker, J. W., & Folger, R. (1987). Beyond negative affectivity: Measuring stress and satisfaction in the workplace. *Journal of Organizational Behavior Management*, *8*(2), 141-158.

Wells, G. L. (1978). Applied Eyewitness-Testimony Research: System Variables and Estimator Variables. *Journal of Personality and Social Psychology*, *36*(12), 1546-1557. DOI: 10.1037/0022-3514.36.12.1546

Werner, N. S., Kühnel, S., & Markowitsch, H. J. (2013). The Neuroscience of Face Processing and Identification in Eyewitnesses and Offenders. *Frontiers in Behavioral Neuroscience*, *7*, 189. DOI: 10.3389/fnbeh.2013.00189

Wühr, P., & Huestegge, L. (2010). The impact of social presence on voluntary and involuntary control of spatial attention. *Social Cognition*, *28*(2), 145-160.

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2009). Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 558.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology*, *18*(5), 459-482.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141-145.

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current biology*, *15*(24), 2256-2262.

Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of applied psychology*, *71*(2), 291.

Yuille, J., Davies, G., Gibling, F., Marxsen, D. (1994). Eyewitness memory of police trainees for realistic role plays. *Journal Of Applied Psychology, 79*(6), 931-936. Doi: 10.1037//0021-9010.79.6.931

Zajonc, R. B. (1965). Social facilitation. Ann Arbor: Research Center for Group Dynamics, Institute for Social Research, University of Michigan.