U.PORTO

MESTRADO EM PSICOLOGIA
PSICOLOGIA DA JUSTIÇA E DA DESVIÂNCIA

Eyewitness memory: how recognition is tested, but not the emotional content, matters.

Júlia Oliveira Cipriano

M

2023

Universidade do Porto
Faculdade de Psicologia e de Ciências da Educação

EYEWITNESS MEMORY: HOW RECOGNITION IS TESTED, BUT NOT THE EMOTIONAL CONTENT, MATTERS

Júlia Oliveira Cipriano

June 2023

Dissertation presented at the Master's in Psychology, specialization in Psychology of Justice and Deviance, Faculty of Psychology and Educational Sciences of the University of Porto, supervised by Dr. ***Alessandra da S. Souza de Carvalho*** (FPCEUP).

LEGAL WARNINGS

The content of this dissertation reflects the perspectives, work and interpretations of the author at the time of delivery. This dissertation may contain inaccuracies, both conceptual and methodological, which may have been identified after its submission. Therefore, any use of its contents must be exercised with caution.

By submitting this dissertation, the author declares that it is the result of her own work, contains original contributions and all sources used are acknowledged, such sources being duly cited in the body of the text and identified in the references section. The author also declares that she does not disclose in this dissertation any content whose reproduction is prohibited by copyright or industrial property rights.

AGRADECIMENTOS

First, I would like to thank my supervisor who guided me through this journey and gave me important insights and tips on having a more scientific view over things. I would also like to thank everyone that took their time to participate in this research.

Thank you, *Arisu*, for being my hype ~~boy~~ gurl and putting up with my awful mood every day, preparing food, sending supportive messages, and dragging me to get some fresh air and nice food that reminded us of home. Thank you, *Ratatouille*, for the immensurable support during this time. You have always been there and even helped me finding materials in other languages because I had this personal goal of having a more diverse background for my research.

Thank you *Fã Clube do Miranha, Amo Vcs*, *Marida* and *Consagrado Madara* for being amazing friends and always providing me with fofoca, memes, pet videos or just generally letting me talk about why I was stressed. That helped keeping me as sane as possible. Thanks *Janis Joplin (dog 2, neném da tia), Jimmy Hendrix (cat, tripé)* and *João Gilberto (dog 3, galeto)* who did nothing and do not even live with me, but your videos of playful fights and trying to catch your own tail made me laugh during such stressful times. Thank you, *Shakira (dog 1, Shaki)*, for appearing in my dreams and keeping my memory of you very much alive.

Shout out to *Woody*, who has always supported and protected me since we were kids, I learnt a lot from you. Even when you would look at me sometimes and think "parece que não há sinal de vida inteligente por aqui", you always said "Os seus problemas são meus também / E isso eu faço por você e mais ninguém / O que eu quero é ver o seu bem / Amigo, estou aqui". No better way to finish this moment by saying "o que quer que aconteça, sempre estaremos juntos. Ao infinito e além.". Muito obrigada *Mamacita e Papacito* por me darem a oportunidade de realizar um sonho, foram anos difíceis para todos nós. Eu literalmente não chegaria aqui se não fosse por esse esforço conjunto. Obrigada Mãe por representar a família e chegar do aeroporto a 10min de me chamarem na formatura – cena digna de filme –, foi uma loucura mas eu amei. Pai, pode relaxar porque sua segunda filha finalmente formou.

Lastly, I want to acknowledge *myself* for surviving these years, in a foreign country far from home, family and childhood but despite your terrible social skills you managed to find people to call family. Past Julia would be shocked with what you accomplished. But future Julia would really appreciate if you started therapy.

Resumo

Condenações injustas devido à identificação errônea do suspeito por testemunhas visuais são um problema no sistema judiciário. Uma das causas propostas deste erro é o conteúdo emocional do crime. A literatura sobre o efeito do conteúdo emocional demonstra resultados inconsistentes, principalmente porque tem concentrando-se na recordação verbal e não no reconhecimento visual do suspeito. Com relação ao reconhecimento visual, um fator crucial é o teste de memória. Tradicionalmente, fotos são apresentadas simultaneamente para identificação. Entretanto, a avaliação de pares de fotos parece eliciar mais identificações corretas. A presente pesquisa visou estudar a) o efeito da emoção negativa vs. neutra no reconhecimento visual, b) o efeito do tipo de teste de reconhecimento (simultâneo vs. pares) e c) se estes fatores interagem. Participantes ($N = 212$) assistiram a um vídeo de um homem e uma mulher interagindo de forma neutra ou negativa. Após ca. 20 minutos, os participantes completaram um teste de reconhecimento visual do homem com apresentação simultânea de seis fotos ou de pares de fotos. O reconhecimento visual não diferiu significativamente entre os participantes que assistiram ao vídeo neutro (M = 78%) ou negativo (M = 77%). Entretanto, houve um efeito do tipo de teste: o teste com pares de fotos produziu uma média de 89,5% identificações corretas contra 65,1% para o teste com todas as fotos apresentadas simultaneamente. Não foi observada interação entre estes fatores. Os resultados sugerem que o tipo de teste parece ser mais decisivo do que o conteúdo emocional na obtenção de identificações corretas.

Palavras-chave: memória, emoção, testemunha visual, lineup, comparação pareada

Abstract

Wrongful convictions due to suspect misidentification by eyewitnesses are a problem within the justice system. One of the proposed causes for this error is the emotional content of the crime. The literature about the effect of the emotional content shows inconsistent findings, mostly because it has focused on verbal recollection rather than the facial recognition of the suspect. Regarding visual recognition, a crucial factor is the type of memory test. Traditionally, pictures are shown simultaneously for identification. However, the comparison of pairs of pictures appears to elicit more correct identifications. The present study aimed to examine: a) the effect of negative vs. neutral emotion in visual recognition, b) the effect of the type of recognition test (simultaneous vs. pairs), and c) if those factors interact. Participants (N = 212) viewed a video of a man and a woman interacting either in a neutral or negative way. After ca. 20 minutes, participants completed a visual recognition test for the male character with either six pictures presented simultaneously or in pairs. Visual recognition did not differ significatively between participants exposed to the neutral (M = 78%) or negative (M = 77%) video. However, there was an effect of test type: the test with pairs of photos produced 89.5% of correct identifications against 65.1% for the simultaneous test with all pictures. There was no significant interaction effect between these factors. Results suggest that the type of test could be more decisive than the emotional content in producing more correct identifications.

Keywords: memory, emotion, eyewitness, lineup, paired comparison

Resumé

Des condamnations injustes dues à l'identification trompeuse des témoins oculaires s'avère problématique pour le système judiciaire. L'une des causes proposées est le contenu émotionnel du crime. La littérature sur l'effet du contenu émotionnel présente des résultats inconsistants, puisque que se concentre sur le souvenir verbal plutôt que sur la reconnaissance visuelle du suspect. Quant à la reconnaissance visuelle, le test de mémoire est crucial. Traditionnellement, les photos sont présentées simultanément pour l'identification. Toutefois, l'évaluation de paires de photos semble défavoriser davantage les bonnes identifications. Cette recherche vise à étudier a) l'effet de l'émotion négative par rapport à l'émotion neutre sur la reconnaissance, b) l'effet du type de test de reconnaissance (simultané ou par paires) et c) si ces facteurs interagissent. Les participants (N = 212) ont regardé une vidéo d'un homme et d'une femme interagissant de manière neutre ou négative. Après ca. 20 minutes, les participants ont effectué un test de reconnaissance visuelle de l'homme avec six images présentées simultanément ou par paires d'images. La reconnaissance visuelle ne différait pas significativement entre les participants exposés à la vidéo neutre (M = 78%) ou négative (M = 77%). Cependant, il y avait un effet de type de test: le test avec des paires d'images a produit 89,5% d'identifications correctes contre 65,1% pour le test simultané avec toutes les images. Il n'y avait pas d'effet d'interaction significatif entre ces facteurs. Les résultats suggèrent que le type de test pourrait être plus décisif que le contenu émotionnel pour produire des identifications plus correctes.

Mots clés: mémoire, émotion, témoin oculaire, lineup, comparaison par paires

Contents

Introduction

When an individual commits a crime and it has been signalized to the State by official representatives, the standard procedure is gathering legal evidence and going to trial. One basic principle of the Justice System is that everyone is not guilty until proven wrong. Therefore, the trial mainly consists of prosecutors trying to present compelling evidence to prove that the person in custody (defendant) is responsible for said crime. The evidence needs to follow standards in the way they were collected, introduced, or analyzed to be considered eligible for use in court, and they can be material (e.g., pictures, audio, clothing) or a person (i.e., a witness).

Although there are different types of witnesses, the most important for the current study is the eyewitness, who is, put simply, an individual that saw firsthand any aspects relevant to the case being judged. Usually, the eyewitness was present at the crime scene or was somewhere that gave them access to relevant events. Another important aspect is that the person should only report facts, and not assumptions, opinions, or any information they did not have access to by themselves. However, unlike other types of evidence, the one provided by eyewitnesses depends on their memories, which are susceptible to several influences such as the emotional content of the witnessed event or even how the event was retrieved, making it less accurate than ideal.

1. Importance of Eyewitnesses' Testimonies in the Legal System

Eyewitness' testimony is considered legally valid as means of providing evidence against someone because, when called to the stand, the individual must give a sworn statement that they will speak the truth and only the truth. Therefore, the judging panel considers the declaration as an objective report of the situation. Yet, a person does not simply share facts, they tell a) the interpretation they made of the situation, and b) what they are able to recover from memory. Consequently, the testimony is subjective.

The implications of this subjectiveness are especially important when it revolves around deciding someone's fate based mostly on a person's testimony. There are several projects worldwide that shine a light on the fallibility of eyewitness testimony. One example is The Innocence Network, which has organizations in the following countries: Argentina,

Australia, Brazil, Canada, Ireland, Israel, Italy, Netherlands, Taiwan, United Kingdom and United States of America (The Innocence Network, 2021). Another example is the Red Inocente, present in Latin America, and that has helped exonerate more than 200 innocents (Red Inocente, 2023).

Dioso-Villa (2015) shared that out of the 71 cases in Australia that were identified as wrongful convictions, 6% were due to misleading eyewitness identification or eyewitness error. Similarly, Dullum (2010 cited in Stridbeck & Magnussen, 2012) suggested that around 10% of the 58 exonerated innocent in Norway were sentenced based on new statements from eyewitnesses (which means cold cases were reopened due to new testimony, which ended up leading to the innocent's conviction). In 2016, the Taiwanese Innocence Project reported that out of the 16 cases they helped exonerate, 5 were due to eyewitnesses' errors (Taiwanese Innocence Project, 2022). In Brazil, out of 90 eyewitnesses' identification cases analyzed, approximately 77% led to preventive detention (with an average duration of nine months) – and when the trial was conducted, the judge confirmed it was an actual misidentification by the eyewitness (Defensoria Pública do Estado do Rio de Janeiro, 2022).

The Innocence Project shared that nearly 70% of the 367 DNA exoneration cases in the USA were related to eyewitness misidentifications (Innocence Project, 2020). In the USA, as of 2021, 47 people were considered innocent after the reopening of the case (The National Registry of Exonerations, 2022) and an analysis of 7 wrongful convictions, exonerated during 2020 and 2021, that led to Death Penalty showed that 2 of them included mistaken eyewitness identification – and a total of 37 since 1970 (Death Penalty Information Center, 2021).

Why do such a high number of wrongful convictions happen? Some of the explanations may lie on the processes that contribute to memory generation, the emotional valence of the events experienced, and how memory was tested.

2. Memory

The process of forming a memory is divided into three stages: encoding, storage, and retrieval. Encoding is the first stage, and it consists of the transformation of the sensory input into a mental representation that can be stored (Campos et al, 2010). Information can be encoded in several ways, and during encoding the original information input may be translated into a different format, such as semantic, visual and acoustic (Mujawar et al.,

2021). For example, the image of a dog may be transformed into the verbal label "dog", with only the latter being retained in memory. This consists on the second stage – storage –, when there is an association between inputs (Campos et al, 2010), and varies depending on the type of information stored, how many items were stored (reflecting limits on storage capacity), its duration, and where it is stored in the brain (Campos et al, 2010; Cowan, 2008). This step directly influences the next stage, which is retrieval. Retrieval reflects the ability to take information outside of storage, either by recalling or recognizing (Campos et al, 2010).

Researchers classify memory systems in terms of their capacity limitation and its duration. A classical distinction is between short-term and long-term memory. Short-term memory (also known as working memory) consists of the information that is maintained for a short period of time in the service of some task goal (Cowan, 2008). Cowan (2008) suggests that short-term memory has a "chunk capacity limit" of about 3-4 items. Long-term memory, conversely, is assumed to store information for an unlimited period, and its capacity is also assumed to be unlimited. Yet, note that although storage is assumed to be unlimited, retrieval from long-term memory depends on a slow and cue-dependent search (Gillund & Shiffrin, 1984), which also can lead to memory failures.

When it comes to eyewitness memory, we are talking about a situation where the person experienced an event that they may not have a specific intention to remember, consisting of a case of incidental memory. Additionally, retrieval of the event may be requested days, weeks or even months later (Sauerland et al., 2016). Therefore, the present study focuses on incidental memory for an event that is retrieved after a delay.

As noted by Głomb (2020), a considerable part of eyewitness identification research includes tasks that require participants to describe what they saw (through written/oral free answer, judging sentences as true or false or other methods). Such tasks require the *recollection* of facts, which could be classified as central or peripheral details – what exactly is meant by these terms, however, diverge between scientists. Heuer and Reisberg (1990) consider as central those details related to the main story that could not be altered without changing the nature of the event (while peripheral details can be changed without interfering with the event). On a different approach, Christianson and Loftus (1991) define central vs. peripheral details according to the spatial setting: peripheral details are related to the background of a scene and central aspects are those happening in the first plan.

However, in the criminal setting it is common for eyewitnesses to be asked to *recognize* a suspect. This requires a comparison between a visually present stimulus (e.g., a

photo) and a mental representation – being a task where people simultaneously receive an input and interprets it (Głomb, 2020). Contrarily, verbal recollection is about the ability to retrieve and verbalize stored information in the absence of the stimuli (Głomb, 2020). Studies show that eyewitness performance varies depending on whether they were asked to verbally recollect or visually recognize the suspect (Houston et al, 2013; Deffenbacher et al, 2004). These two types of memory tests may point to different conclusions about one's memory accuracy.

2.1. Memory susceptibility

In his book *On the witness stand,* Münsterberg (1908) reports two studies about the relationship between emotional events and memory recall. The first one was conducted in a classroom during the lecture of Professor von Liszt, where actors pretended to have a fight and, during the physical encounter, the revolver one was holding goes off. Students were asked to write a report about the witnessed event. There were around 15% more mistakes in the report of the most emotional part of the event than in other parts. The second study occurred during a meeting between professionals (physicians, psychologists, and jurists). During the meeting, two people enter the room and start having an argument. One of them was holding a revolver and, once the other individual fell to the ground, there was a gunshot sound and both of them got up and ran. None of the people that witnessed the scene knew this was staged, except for the President, who later asked everybody to write reports on what they saw. Analyses of the reports showed that 39 (out of 40) had more than 20% omissions and in ¼ of the papers, more than 10% of what was written was false. This study suggests that even people trained to observe events are susceptible to memory mistakes.

Mittal et al. (2013) discuss how complex yet influenceable an individual's memory is, defining two categories of factors that affect eyewitness' accuracy. One category refers to the features of the situation itself, such as the location, time, how violent it was, and the relevance of the event for the person witnessing. Here, it's also important to mention one phenomenon known as Weapon Focus: it is the tendency to spend more time looking at the weapon/unusual object, if there's one present (Loftus et al., 1987) compared to common/expected object and even the suspect. Accordingly, information retained about a criminal might not be precise, leading to a wrongful identification.

The second category is related to intrapersonal characteristics such as implicit bias (Trawiński et al., 2021), race (Wong et al., 2020; Yaros et al., 2019), stereotypes (Osborne & Davies, 2012), and emotional arousal – either enhancing memory (Saraiva et al., 2021; Mittal et al., 2013; Hulse et al., 2007) or impairing it (Ghani & Ning, 2016; Loftus et al., 1987). When talking about emotional influence, it is necessary to distinguish between positive and negative emotions, as some studies (e.g., Talarico et al., 2009; Forgas et al., 2005) suggest that positive affect influences memory positively whereas negative affect impairs it (leading to the incorporation of misleading information).

3. Emotion

Although emotion is part of the human nature and it is present in daily life, there is no consensus in the literature on its definition (Izard, 2010). That is why it is suggested that studies should define what the authors consider as emotion (Izard, 2010; Houston et al, 2013; Głomb 2022). The present study follows the definition used by Izard (2010):

> *"Emotion consists of neural circuits (that are at least partially dedicated), response systems, and a feeling state/process that motivates and organizes cognition and action. Emotion also provides information to the person experiencing it, and may include antecedent cognitive appraisals and ongoing cognition including an interpretation of its feeling state, expressions or social-communicative signals, and may motivate approach or avoidant behavior, exercise control/regulation of responses, and be social or relational in nature."* (p. 367)

Emotions can be classified according to several models/theories. One example is the discrete model, which considers that individuals are born with basic emotions (like happiness, sadness, and anger). Yet there is no consensus in literature as to which emotions are considered the basic ones, as it is greatly influenced by the sociocultural context (Harmon-Jones et al, 2017). A second model is the dimensional approach that considers that emotional experiences vary in continuous dimensions such as valence and arousal. The arousal factor refers to the physiological changes the emotion elicits (reflecting the activation of sympathetic nervous system) being a subjective experience classified within a spectrum (Harmon-Jones et al, 2017). As for the affective valence, it is related to how an individual

analyzes their experience, classifying it as positive (or of pleasure) or negative (or of displeasure) (Harmon-Jones et al, 2017). There are also models that mix both theories (dimensional and discrete approaches). In the present study, we will ask participants to rate the intensity of their emotions using discrete categories (e.g., sadness, happiness, disgust).

3.1. The Impact of Emotions on Eyewitness Testimony

During the review of the empirical literature, a mixed pattern of results was observed regarding the impact of negative emotion on long-term retention of memory details. Some studies observed that negative emotion enhanced memory and others observed that it impaired it, making it necessary to analyze the type and intensity of emotion, what type of details were better/worse recalled, as well as the type of memory test.

Christianson and Loftus (1991) conducted a study using as stimuli 15 colored slides that varied only on what was depicted on the eighth slide, creating three different situations. In the *neutral* condition, said slide showed a woman bicycling in front of a car on the street; in the *negative* condition, the woman was seen lying near her bicycle, head bleeding, in front of a car on the street; lastly, in order to study if the effect on memory wasn't related to the scenario not being common/expected, there was a condition called *unusual*, where the woman was seen at the same spot, walking in front of the car but carrying the bicycle on her shoulders.

To test memory, participants were shown again the same slides, but the eighth slide didn't show the woman nor the car. Participants were asked to write down what was missing, including colors, type of objects and forms. Next, there was a recognition task where participants were presented with four similar slides, one being the one shown on the study phase and the other three being fillers (changing the color or the woman's coat and/or color of the car) and asked to choose the one they saw previously. Results showed that participants in the negative emotion condition were a) worse both in recognizing and recalling peripheral details (e.g., realizing the car wasn't there) and b) better at recollecting central details (e.g., realizing the woman wasn't there). There were no significant differences between groups in recognizing central details.

Dutton and Carroll (2001) studied the impact of negative emotion using as stimuli video excerpts and written passages about three movies. The written passages were used to induce different levels of emotion: low, medium, and high arousal. A week later, participants

answered questions about central and peripheral details of the videos. The high arousal condition impaired recollection of central details (character's description) when compared to the low and medium arousal conditions. There was with no significant difference between the groups for peripheral details. Hence, unlike Christianson and Loftus' (1991) study, there was no difference between groups for peripheral aspects, whereas memory for central details was impaired.

Mittal et al (2013) used as stimuli a video depicting a bank robbery and manipulated the negative emotion through pre narrations about the crime scene – therefore having a group of low arousal and another with higher arousal. The recollection task consisted of participants answering if a certain detail was present or not in the video. Immediately after the video, the higher arousal group remembered more correct details about central aspects when compared to their lower arousal counterparts – while there were no significant differences for peripheral details. After a 1-week delay, however, there were no significant difference between groups.

In their meta-analytic review about the impact of stress on recollection and recognition tasks in eyewitness memory, Deffenbacher and colleagues (2004) found considerable evidence to support the hypothesis that stress hinders accuracy performance on both tasks. Additionally, the authors suggest that although the direction of the effect was similar, performance on the facial recall was independent from the facial recognition. This means someone could have a satisfactory result in one but not achieve the same result in the other. For this review, 16 studies published between 1974 and 1997 were included, with a total of 27 face recognition tasks, 36 details recall tasks and 1727 participants involved.

Marr et al. (2021) investigated the effect of stress on face and word recognition. To induce stress, they used the Maastricht Acute Stress Test and checked for physiological signs and negative affect scores. Stress participants were presented with the stress inducing task while viewing face pictures or words and later reported their emotion. The control group performed non-stressing control tasks while studying the information. After 24-26 hours, they were given the picture or word recognition test. There was no effect of acute stress on recognition of pictures or words.

Saraiva et al. (2021) divided participants in three groups according to the intensity of the emotion: the control group watched a video depicting a random man chatting with a woman; the lower emotion video showed the man robbing the woman; lastly, the higher emotion video depicted the robbery followed by murder. After answering an anxiety test and doing a filler task, participants indicated the veracity of a list of statements about the video,

as well as their confidence on their response (between 0% and 100%). Finally, participants identified the suspect in a simultaneous lineup of six pictures. There were no significant differences between the groups neither for accuracy of verbal recall nor on correct suspect identification. In a second experiment, Saraiva et al. (2021) manipulated the retention interval: instead of doing the task about video recollection and suspect identification five minutes after watching the videos, participants were tested after a delay of 7, 14, or 21 days. Again, there were no significant difference between groups in the verbal recollection and suspect recognition tests.

Snow (2022) examined the influence of negative emotion on eyewitness memory, both for recollection and recognition. After participants watched either a negative or a neutral video and rated their emotions, memory was tested either immediately (delay of around 2 minutes) or 1 week after. In the recollection (verbal recall) task, participants were requested to type what they could remember about the video, either freely or with more detailed prompts derived from the Cognitive Interview protocol. As for the recognition task, participants were provided with a simultaneous lineup. The negative emotion group showed enhanced memory for central actions (but not face details) but there were no significant differences between groups in the lineup identification. Concerning the delay, the negative emotion group was less likely to identify the suspect after one week.

Houston et al (2013) also investigated the effect of negative emotion on eyewitness testimony. To induce emotions, participants were shown a neutral or negative video depicting the interaction of an old woman and a younger man. After seeing the video, participants completed an emotional rating scale to assess their emotional state, and then a filler (unrelated) task for ca. 20 minutes. Finally, participants were asked to recall verbally the events (Experiments 1 and 2) and to recognize the man in a six-photo simultaneous lineup (Experiment 2). Participants in the negative group verbally recalled more details about the man than the neutral group, although the accuracy of the descriptions was similar between groups. However, participants in the negative group were worse at correctly identifying the man when he was present in the lineup.

In sum, studies including recollection tasks (verbal recall) show mixed results: some suggest that negative emotion – or higher level of such emotion – impairs memory while others suggest that it enhances it. However, such results should be interpretated carefully, as eyewitness memory can be divided in central and peripheral aspects, and performance varies according to such categories. When it comes to recognition tasks, the literature shows either that negative emotion tends to impair one's ability to recognize a suspect or that there is no

difference to neutral situations. More studies are therefore needed to clarify the impact of negative emotions on suspect recognition.

4. The Role of the Suspect Identification Test

Traditionally, memory for the suspect's face is assessed through a visual recognition test that can use pictures, live lineups, and, more recently, videos (Kemp et al, 2001; Davis et al, 2015). The use of pictures is common in the laboratory, and it is also frequently used by the police. The test can be implemented in two ways: through a *simultaneous lineup*, where police show several faces (usually six) at the same time, or through a *sequential lineup*, where the pictures are shown one-by-one (Gepshtein et al, 2020). In the laboratory, the lineups can be manipulated to include the picture of the suspect and of people that are known to be innocent (also called fillers). This is known as the target present lineup. The lineup can also be constructed without the picture of the suspect (target absent lineups). Although it is known that police officers can influence the eyewitness by the way they behave or talk (Innocence Project, 2020; Brewer & Wells, 2009; Douglass & McQuiston-Surrett, 2006), fewer studies have investigated the effect of the lineup itself on performance.

Akan et al (2021) investigated how the lineup size can affect identification, and their study suggest that the number of fillers doesn't impact discriminability. Brewer and Doyle (2021) suggested a more radical change: instead of picking a person, the witness should share their confidence level about the similarity between their memory and each of the portraits. Although there is compelling theorical reason behind this logic, the main problem with this suggestion is how open are criminal justice professionals to the method given that it creates an opportunity for defendant's lawyers to question the identification. Imagine that an eyewitness is 90% confident, the remaining 10% are enough to create reasonable doubt, and hence to discredit the testimony.

Gepshtein et al. (2020) showcased the potential of a procedure known as perceptual scaling and ROC (receiver operating characteristics) analysis in the criminal field. This method proposes to identify central tendencies when recollecting characteristics of a suspect by making the person select repeatedly between two faces (hereafter referred to as a pair evaluation procedure). In each round, the eyewitness selects from the pair the picture of the

individual that more closely resembles the image in their memory. By doing this, we are able to gather important information from each option (e.g., its similarity), instead of discarding it after the selection of one individual.

Another important benefit of the method is that it lessens the burden on the person being inquired, as they do not need to say "it is this person" or "the person is not here" – which is a dichotomic decision. As the person does not have to choose a single picture, a correct identification happens when the picture of the suspect is selected more times when asked "which is closer to what you remember" when compared to the other pictures. If the suspect is not in the combination presented, each picture should have a similar probability of being selected. This procedure allows one to measure which face is more likely to match the one in memory while at the same time avoiding the problem of decision criterion that arises in all-or-none lineups.

5. Present study

The present research examined whether negative emotion when witnessing a crime scene affects memory retrieval for facial information, reducing the accuracy of suspect identification. Additionally, we aimed to determine if the effect of emotion on the identification task depends on the method of testing memory, which may lead to more biased judgments. More specifically, our goal was to answer the following questions:

a) Does negative emotion hinder the suspect identification?
b) Does the pair evaluation produce more correct identifications than a traditional simultaneous lineup?
c) Does the pair evaluation test mitigate the impact of negative emotion on suspect identification?

To answer those questions, we created four groups of participants that differed regarding the emotional content of the witnessed event (neutral vs. negative) and the type of memory test (traditional simultaneous lineup vs. pair evaluation). The current study was modelled on the experiment conducted by Houston et al. (2013) but bears some modifications in the methods. The first modification was related to the self-assessment of the emotional state of the participants after presentation of the video with the critical event.

Houston et al. (2013) used a 4-point scale to measure emotions. Here we employed an 11-point scale, as a higher number of points can increase the sensibility of the measure, reducing skewness and kurtosis (Głomb, 2022). The second one is the absence of a memory test requiring the verbal recall of the events. We removed this phase because implementation of verbal recall before the suspect identification can reduce face recognition due to the Verbal Overshadowing Effect, which refers to one's decreased ability to recognize a face previously seen when they are asked to make a description/are provided with a description of a face they had previously seen (Dodson et al., 1997). Finally, Houston et al. (2013) used a simultaneous lineup for suspect identification with a version with the target present and a version with the target absent. Here, we only included the target present version, because no differences were observed between groups in the target absent lineup in the previous study. Finally, we tested memory using either the simultaneous lineup or the pair evaluation method proposed by Gepshtein et al (2020).

Our first hypothesis is that negative emotions impairs memory. Accordingly, we expect that participants in the negative group will have poorer accuracy when recognizing the suspect seen on the video compared to the neutral group. Conversely, the null hypothesis would be that there is no impact of negative emotion on recognition. Here, the term "negative emotional distress" follows the classification suggested by Głomb (2022), which refers to affective states classified as anxiety, irritation, fright, etc. To measure the intensity of said emotions, a self-report scale was used, presenting discrete types of emotion that were evaluated regarding their intensity in the present moment.

One explanation for the reduced face recognition after a negative event could lie on the procedure used to identify the suspects, as sequential and simultaneous methods have known limitations (Palmer & Brewer, 2012; Steblay et al, 2001). Therefore, the second goal of the present study was to compare the methods of suspect identification, and how they interact with the emotional context. Our second hypothesis was that the pair evaluation method improves identification overall compared to the simultaneous lineup as proposed by Gepshtein et al (2020). Our third and final hypothesis is that this method may reduce the deleterious impact of negative emotion on suspect identification: in other words, the negative emotion group may no longer be worse than the neutral emotion group when the memory test requires the evaluation of all pairs of suspect photos.

Method

1. Participants

Participants were recruited from three Psychology course lectures (Psychology of Attention and Perception, Psychology of Thinking and Language, and Neuropsychology) as well as through social media. Inclusion criteria were being of legal age (18 or older), being Portuguese speaker, and having normal or corrected vision. The study was approved by the Faculty of Psychology and Educational Sciences of the University of Porto Ethics Committee, document number Ref.ª 2022/10-05.

Data of 214 participations were received, but two had to be excluded from the study as they consisted of repeated answers – leading to a final sample of 212 participants (153 women, 47 men, 6 identify as non-binary and 1 chose "other"; 5 participants didn't answer the demographic questionnaire), with ages varying from 18 to 65 (M = 24.97, SD = 9.41). Regarding nationality, the majority was Portuguese (64.73%), followed by Brazilian (29.47%), and 5.8% "other". Data about race/ethnicity was collected, but lost due to an coding error. Regarding educational level, 62.32% completed high school, 34.3% bachelor degree, 2.41% a master degree, and 0.97% chose "other". Out of the total sample, 150 were students from the Faculty of Psychology and Educational Sciences of the University of Porto and were rewarded with course credits. Participants were randomly assigned to one of four groups: Negative – SIM (n = 51), Negative – PAIR (n = 51), Neutral – SIM (n = 55), and Neutral – PAIR (n = 55).

2. Materials

2.1. Videos

We presented participants with one of two possible videos. The videos used in the study consisted of edited versions of the first case shown on episode 8 (Bankrupt Heart) of

season 3 of the crime series *Behind Mansion Walls*. Since the episode makes a reconstruction of a real-life murder, first they dramatize the day-to-day life of a couple and, as the episode develops, there is a dramatization of the actual events related to a murder scene.

We edited the video to create two versions of the story: a neutral and a negative one. The *neutral video* shows a couple talking casually while the woman makes a painting. Later, the man is shown writing emails in his computer alone in the room (with a quick preview of the email's contents). Then, the couple is together by the stairs near the house's entrance hall, having an argument. After a few seconds, the woman goes upstairs, and the man is seen calling for her. Later, she comes back downstairs with a luggage, hands him a list, he signs it, they say goodbye, and the man opens the door so the woman can leave, closing it right after she leaves, while reading the list.

The *negative video* starts with the same couple having a slight discussion (instead of small talk) while the woman is painting. Then, it shows the same scene where the man is writing an email. Next it changes to the couple together by the stairs having an argument, and the woman goes upstairs. As she returns downstairs, instead of the peaceful walk shown in the neutral video, she walks angrily, waving a piece of paper (without the luggage), having yet another fight with the man. She hands him the paper; he reads it and after a few seconds she takes it from his hand. At this point, the man clearly gets mad, grabs a nearby wrench and hits the woman. She falls to the ground, where he hits her a few more times. After a quick break (camera fades out to the chandelier), the now bloodied wrench is shown on the floor, which the man picks up. Then, the scene cuts to the man dragging a plastic bag to the door, leaving a trace of blood behind.

To ensure both film clips were as similar as possible, they were edited to last 100 seconds, with each showing the face of the man clearly for 21 seconds (hereafter the suspect). Although the audio could have helped inducing the emotions, it was muted since the native language of the participants of the present study (Portuguese) is different from the language spoken in the video (English). Therefore, this measure secures the subjects will pay attention to the actions, characters and story unfolding without secondary conflicts/interferences (such as a different language or subtitles).

2.2. Emotional Rating Scale

A self-report 11-point scale (being “0” not at all, “5” the middle point and “10” very much so) was used to request a self-assessment of the intensity of the emotions participants felt when watching the videos. Participants were presented with a list of 12 discrete emotions: sadness, sympathy, relief, disgust, irritation, anger, outrage, annoyance, happiness, upset, anxiety, and fright. Relief, sympathy, and happiness were considered as positive emotions, the remaining ones as negative emotions.

2.3. Filler Task

After the emotional rating, participants completed a filler task about visual memory (aka a continuous color reproduction task) for ca. 20 minutes. The task required the memorization of six colored squares for a brief period, and their subsequent reproduction using a color wheel. Participants completed two practice trials and 60 test trials of this task. This task served to create a delay before the test of the memory for the face of the individual shown in the video, assuring therefore that participants had to retrieve the face from their visual episodic long-term memory. Performance on the filler task will not be discussed here.

2.4. Photographic Lineups

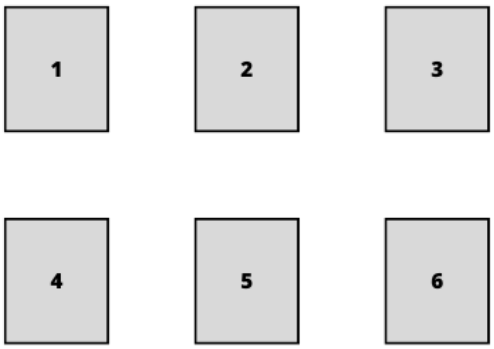
Participants were exposed to one of two possible types of memory test. Half of the participants in the neutral and negative groups were exposed to a simultaneous lineup. The simultaneous lineup consisted of six colored face pictures shown simultaneously, one being the actual perpetrator and the others, fillers with similar features and facial expressions as the person portrayed in the video (as recommended by US Technical Working Group for Eyewitness Evidence, 1999). All faces were presented against a gray background (to minimize possibilities of unintentional clues given by the background). This type of lineup is illustrated in Figure 1. The present study only used a target present lineup. The photo of the suspect could appear in any of the six positions, which was randomly determined for each participant.

The other half of the participants was exposed to a paired comparison test. Participants were exposed to the images of two faces and had to indicate which of the two was more similar to the male character they saw in the video. Participants were exposed to all six photos as in the traditional lineup, but always in pairs (in all possible combinations as

illustrated in Figure 2). In this method, a correct "identification" means that the picture showing the suspect was chosen more times as the one most similar to the participant's memory, meaning it was top ranked.

Figure 1

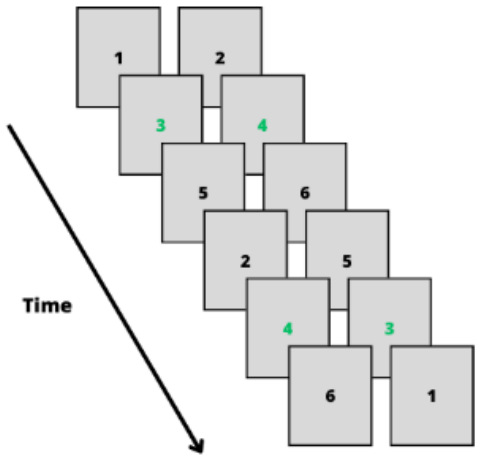
Illustration of the Simultaneous Lineup.



Note. This figure exemplifies the simultaneous lineup in which six face pictures (here shown as numbers) were presented simultaneously. Participants had to click on one of the images to indicate which individual they recognize as the man in the video.

Figure 2

Illustration of the Paired Comparison Test.



Note. This figure exemplifies the task provided to subjects, where they were shown pairs of face pictures (here shown as numbers) and had to choose which one resembled the most their memory of the suspect. Colored numbers exemplify a pair repetition but interchanging places. Adapted from Gepshtein et al. (2020).

3. Procedure

The experimental task was programmed using lab.js (https://lab.js.org/), an open-source experiment builder. This software allows researchers to design their experiment without having to code, but also to use JavaScript, HTML and CSS to customize the study (Henninger et al., 2022). The present study was prepared using a mix of the software's pre-available content supplemented with HTML programming. It was launched on MindProbe (https://www.mindprobe.eu/), running JATOS (https://www.jatos.org/). Participants completed the experiment online.

Participants were told that the purpose of the experiment was to test their visual abilities. Participants read an informed consent form before the start of the study. They were warned that the study may include the depiction of a video with content that could be upsetting for some viewers, as well as warnings about violence and blood. After consenting to the study, the experiment started with the presentation of one of the videos (either neutral or negative). Half of the participants were assigned to the neutral and the other half to the

negative video condition. At the end of the video, participants completed the emotional rating scale to assess how the video altered their emotional state.

Next, participants completed the visual memory filler task for ca. 20 min. Finally, participants received a surprise test of their memory of the face of the man in the video (suspect). Half of the participants in the neutral and the negative groups were assigned to the simultaneous lineup test: they were simultaneously shown six faces and they had to pick the one that they believe was the man in the video. There was only one opportunity to answer. The other half of the participants in the neutral and negative groups responded to the pair evaluation procedure. They were shown pairs of pictures drawn from the six pictures shown in the simultaneous condition. There were 15 possible combinations of the six pictures and the procedure was repeated three times for each pair (for a total of 60 responses).

After finishing all the steps mentioned before, participants completed a sociodemographic questionnaire, asking their age (in years), what gender they identify as (with options "female", "male", "non binary" and "other"), their nationality (options "Brazilian", "Portuguese" and "other"), country of residence ("Brazil", "Portugal" and "other"), race/ethnicity ("white", "latino/a", "black", "Asian", "other"), what is their highest educational level ("not applied", "basic education", "intermediate education" and advanced education – divided in "bachelor's", "master's" and "doctoral". There was also the option "other") and, lastly, their current professional situation (with options "only studies", "only works", "works and studies", "doesn't' study nor work" and "other").

Finally, participants were directed to a debriefing page, stating that the real purpose of the experiment was to study the effect of negative emotion on eyewitness memory as well as the impact of the type of memory test on the accuracy of face identification. They were also thanked for the participation on the experiment and shown, once again, the contact information in case they wanted to ask questions, provide feedback and/or make complaints/report harm due to the experimental manipulation.

4. Design and Analyses

The experiment involved a 2 x 2 between-subjects design consisting of the manipulation of the emotion content of the video (neutral vs. negative) and the type of

memory test (simultaneous, SIM vs. pair evaluation, PAIR). We performed two analyses to answer our main research questions. First, we evaluated if the manipulation of emotional content of the videos was successful by examining the emotional intensity ratings for negative and positive valenced emotions between the groups exposed to the negative vs. neutral content videos. For this, we submitted the average emotional ratings in each of these valence conditions to a mixed ANOVA with valence as a within-subject factor, and video content and test-type as between-subject variables.

Second, we assessed the effect of emotion and type of test on general accuracy of suspect recognition. Responses were classified as correct or incorrect (meaning the subjects identified the male character shown on the stimuli video or identified one of the filler pictures, respectively). For the SIM test, this resulted in a value of 0 (incorrect) or 1 (correct) for each participant, given that they only made one choice. For the PAIR test, this resulted in a proportion between 0 and 1, since participants had the opportunity to select the correct suspect in 15 pairs (in the remaining 45 trials, the suspect was not present, and this data will not be analyzed here). The mean accuracy in the suspect recognition was analyzed in a factorial between-subjects ANOVA having video content and test-type as factors.

All analyses were performed in R (R Core Team, 2022), and RStudio, version 4.2.2 (2022-10-31 ucrt) (R Core Team, 2022). We used the package afex (Singmann et al., 2023) and emmeans (Lenth, 2023) to perform the statistical analyses. We plotted the data using ggplot (Wickham, 2016). As a measure of effect-size, we report generalized eta-square (η_G^2).

Results

1. Emotional Ratings

After presentation of the video, participants rated their emotional state. We used their ratings to assess if the emotional content of the video was effective to induce negative and neutral emotions. As mentioned before, the emotions in the emotional rating scale were aggregated and divided according to their valence (positive or negative). The neutral and

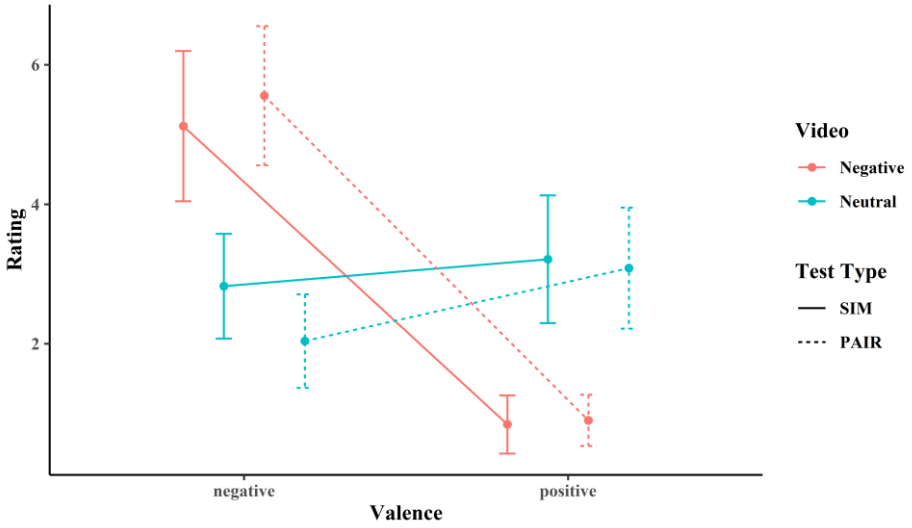
negative groups' ratings were compared through a mixed factorial ANOVA, with "valence" as the within variable and "video" and "test type" as the between subjects factors.

Figure 3 presents the mean ratings for the positive and negative valence emotions for the four groups that vary with regards to the emotional content of the video and the test-type. Note that the rating was performed before exposition to the memory test, and hence this variable is not expected to have an effect on the emotional ratings. There was a significant effect of valence, $F(1,208) = 128.4$, $p < .001$, $\eta_G^2 = .18$, and an interaction between video and valence, $F(1,208) = 245.34$, $p < .001$, $\eta_G^2 = .29$. As expected there was no significant effect of test-type, $F(1,208) = 0.21$, $p = .65$, $\eta_G^2 < .001$. Follow-up tests showed that the negative video groups (considered together) rated the negative emotions higher ($M = 5.34$, $SE = 0.26$, 95% CI [4.83, 5.85]) than the positive emotions ($M = 0.87$, $SE = 0.1$, 95% CI [0.68, 1.07]), $t(100) = 17.15$, $p < 0.0001$. Regarding the neutral video group, positive emotions were rated higher ($M = 3.15$, $SE = 0.22$, 95% CI [2.71, 3.59]) than the negative emotions ($M = 2.43$, $SE = 0.18$, 95% CI [2.08, 2.78]), $t(108) = -3.45$, $p = 0.0008$.

When comparing the groups in each valence condition, it was observed that the negative groups had significantly higher ratings for negative-valenced emotions than the neutral groups, $t(208) = 9.388$, $p < 0.0001$. For the positive-valenced emotions, however, the negative group reported significantly lower ratings compared to the neutral group, $t(208) = -9.12$, $p < 0.0001$.

Figure 3

Emotional Intensity Ratings for the Negative and Positive Valenced Emotions for the Four Groups that Differed According to Emotionality of the Video and the Memory Test Type.



Note. Error-bars represent 95% within-subjects confidence intervals.

The rating's mean and standard deviation for each emotion category, analyzed by video content (neutral vs. negative), are presented in Table 1. Figure 4 depicts the distribution of the ratings for each emotion considering video content (neutral and negative) and test type (SIM and PAIR).

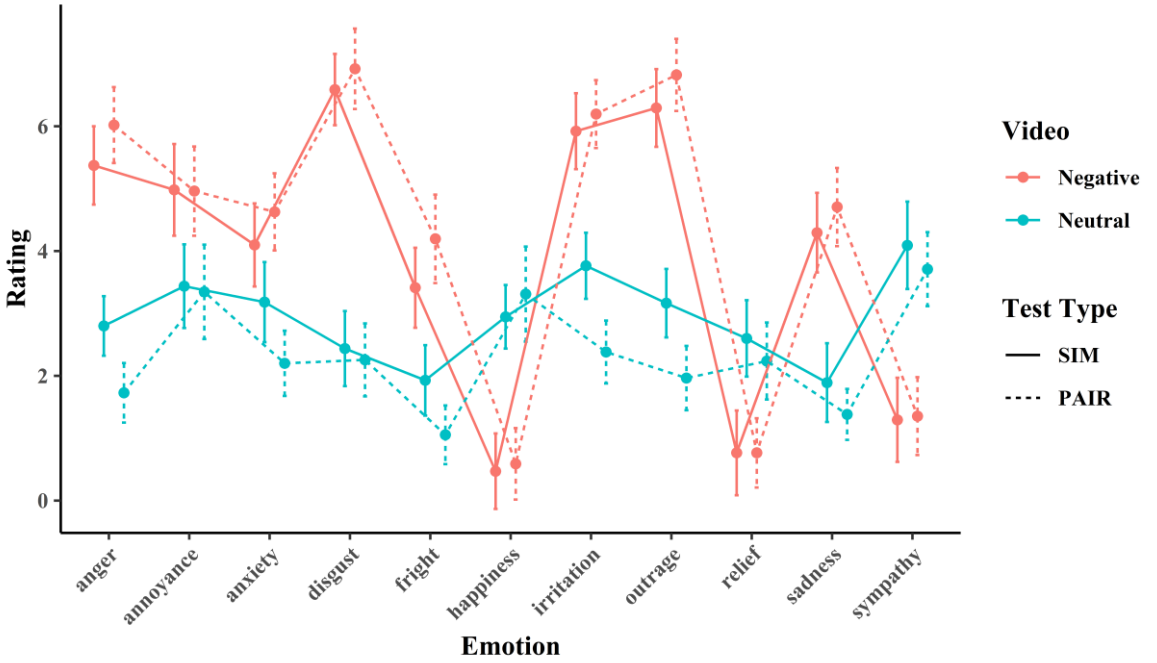
Table 1

Average Emotional Ratings for each of the Emotions in the Scale for the Neutral and Negative Groups. Note that Data of the Groups with the two Types of Tests were Collapsed.

Emotion	**Neutral Video**		**Negative Video**	
	M	*(SD)*	*M*	*(SD)*
Anger	2.27	(1.77)	5.7	(2.2)
Annoyance	3.4	(2.64)	4.97	(2.59)
Anxiety	2.69	(2.16)	4.37	(2.29)
Disgust	2.35	(2.19)	6.76	(2.16)
Fright	1.49	(1.92)	3.81	(2.4)
Irritation	3.07	(1.91)	6.06	(2.05)
Outrage	2.56	(1.97)	6.56	(2.14)
Sadness	1.64	(1.92)	4.5	(2.25)
Mean	2.43	2.06	5.34	2.26
Sympathy	3.9	(2.39)	1.32	(2.31)
Happiness	3.13	(2.35)	0.53	(2.1)
Relief	2.42	(2.28)	0.76	(2.19)
Mean	3.15	2.34	0.87	2.2

Figure 4

Average Emotional Ratings for each of the Emotions in the Scale for the Neutral and Negative Groups, divided by Test Type.



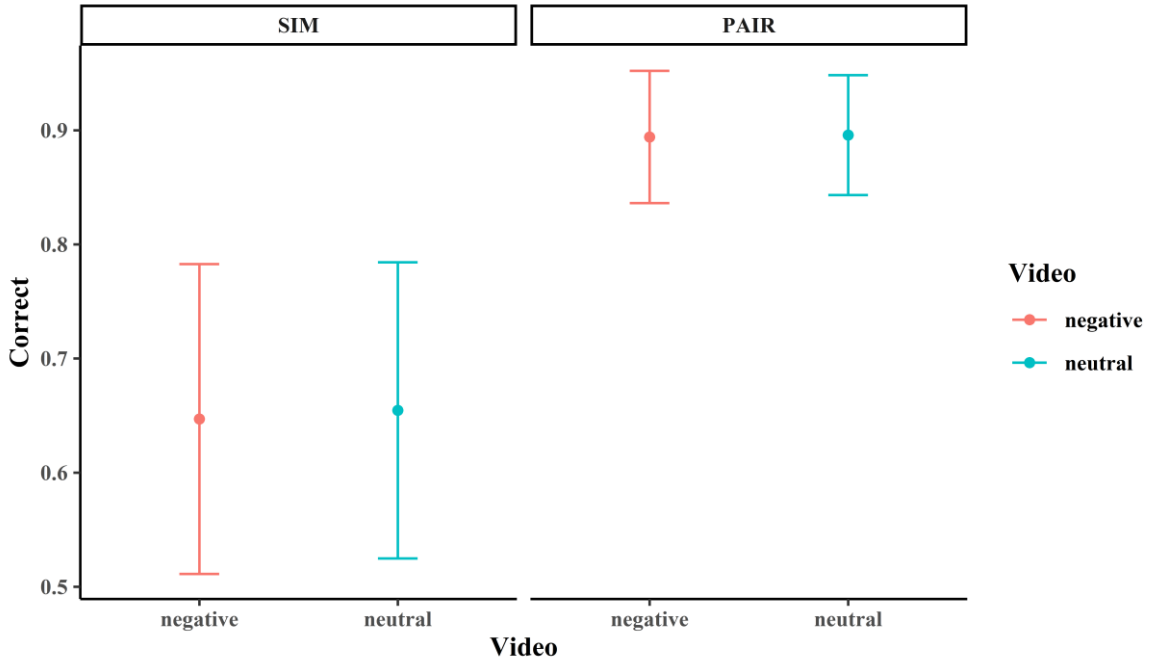
Note. Error-bars represent 95% within-subjects confidence intervals.

2. Recognition Performance

To analyze the effect of emotional content and test type in recognition performance, a factorial ANOVA was implemented with correct identifications as the dependent variable and the video (negative or neutral) and type of test (SIM or PAIR) as between-subjects factors. The sample consisted of 212 participants. The dependent variable consisted of the proportion of correct identifications in the SIM lineup, and the average proportion of selections of the suspect when compared to any of the other pictures in the PAIR lineup. Figure 5 represents the average proportion of correct identifications as a function of the emotional content of the video and the type of test.

Figure 5

Average Correct Recognitions based on Test Type and Emotional Content.



Note. Error-bars represent 95% within-subjects confidence intervals.

There was no main effect of the video condition, $F(1,208) = .01$, $p = .93$, $\eta_G^2 < .001$. The average performance of the neutral groups ($M = 0.78$; $SE = 0.05$, 95% CI [0.68, 0.87]) was similar to the average performance of the negative groups ($M = 0.77$; $SE = 0.05$, 95% CI [0.67, 0.88]). We found a main effect of the type of recognition test, $F(1,208) = 23.22$, $p < .001$, $\eta_G^2 = .1$, because the SIM groups had a lower average proportion of correct responses ($M = 0.65$; $SE = 0.5$, 95% CI [0.56, 0.75]) than the PAIR groups ($M = 0.9$; $SE = 0.5$, 95% CI [0.8, 1]). However, there was no effect of the interaction between video and test type, $F(1,208) = 0$, $p = .95$, $\eta_G^2 < .001$.

Discussion

The current study aimed to answer three main questions: a) does negative emotion interferes with facial recognition performance, b) does the pair evaluation procedure elicits more correct identifications than the simultaneous lineup, and c) does the pair evaluation mitigates the effects of negative emotion on face recognition. To answer those questions, first, it was necessary to induce the neutral and negative emotions on the participants while they observed the suspect. To attain this goal, we created two videos that depicted either neutral interactions between a couple, or negative interactions that culminated in a crime. The group that viewed the crime video (negative content) reported higher values of negative emotions than the neutral group. These results indicate that the crime video elicited the purported negative emotions in the participants.

Contrary to our first hypothesis, no effect of negative emotion was observed in the suspect recognition task. These results contradict the ones found on the study of Houston et al. (2013) that served as the basis for the current study. Houston and colleagues (2013) observed worse memory for the negative group (27% correct identifications) than the neutral group (40% of correct identifications), although the effect was small (Cramer's V = .24). The present findings are in line with more recent experiments that suggest that negative emotions have no influence on recognition accuracy – neither enhancing nor impairing it (e.g., Snow, 2022; Saraiva et al., 2021).

In the Houston et al. (2013) study, emotion reports were 1.28 (SD = 0.73) for the negative group and 0.20 (SD = 0.25) for the neutral group. However, one should bear in mind that the authors used a 4-point scale. Converting their results to the 11-point scale used in the present study, the negative group reported an average of 4.27 (SD = 2.43), which is slightly lower that then average observed in the present study 5.34 (SE = 0.22), while the neutral group reported an average of 0.67 (SD = 0.83) – compared to M = 2.43 (SE = 0.22) in the present study. Values were converted using function rescale of package *scales* (Wickham & Seidel, 2022). Thus, both our negative group and the one in Houston et al. reported similar levels of negative-valence emotions. Hence it seems unlikely that a lower level of induced negative emotion could explain our findings.

Our second hypothesis, namely that the pair evaluation method improves identification overall compared to the simultaneous lineup, was supported corroborating the original study (Gepshtein et al., 2020). The current study has a larger sample than the original

study (95 vs. 212 participants) which provides more robust evidence for the advantage of this method.

The advantage of this method could translate into different benefits for several parts inside the court room. For the jury/judge, this method brings a quantitative overview about how many times the person was chosen among the total number of visualizations, which could point to a stronger testimony from the eyewitness. Thus, it could lead to a possible decrease of the number of cases where eyewitness misidentification contributed to a wrongful conviction. For an actual innocent defendant, it could prevent a violation of their human rights and consequent social and personal impacts of being incarcerated. Besides, it also brings benefits to the State, as it lessens the financial troubles derived from wrongful convictions, considering there is a price for maintaining people incarcerated but there are also compensations for wrongfully convicted people. For police officers, this method dismisses the need of conducting target absent lineups as the trials that do not show the suspect can also be computed – therefore comparing the number of times each picture was chosen. Lastly, as for the eyewitnesses themselves, it could decrease the burden felt while doing such task, as they never really point to a specific person – although police officers are supposed to tell them an identification is not mandatory, some eyewitness might still feel the need to.

With regards to our third question, namely whether the PAIR procedure could mitigate the deleterious effect of negative emotions on recognition, we found no evidence for an interaction between emotion and the test procedure. Yet, given that we could not find any effect of the negative emotion on suspect identification, this hypothesis remained untested.

Considering that video stimuli may have a number of confounds (such as having different scenarios/plots used for each experimental group, leading to one video possibly being more interesting than the other; Głomb, 2022), special care was taken when preparing the video material for the current study. To prevent such problem, instead of using videos from different sources for each group, the exact same scenes/aspects were used for both conditions, varying only the necessary contents for emotion inducing purposes. Other aspects considered were having the same exact total video duration, as well as same duration of exposition to the face of the suspect, and minimizing distracting factors within the stimuli (e.g., muting the sound due to discrepancy in language). Such measures allowed a more precise comparison between groups (neutral vs negative emotion) leading to a more consistent result and interpretation. Thus, our results align with a growing number of studies

that found no significant effects of emotion on eyewitness memory, suggesting that there is no impact of induced emotion on facial recognition.

This study has some limitations. The first issue refers to how emotions were evaluated: instead of a hybrid approach with both self-report and physiological measures, it was based solely on self-report. This design choice replicates the one employed in the study of Houston et al. (2013). Self-report is subject to social demand, and it is unclear if participants might have reported the emotions they thought they should have felt when seeing the video or their real experience. This limitation, however, applies both to our study and the one of Houston et al. (2013) and it cannot explain the different results observed across these studies. Another potential limitation of our study (and also of Houston et al.) was that we did not access the emotional state of the participants prior to the presentation of the videos to more critically determine whether the video content changed their emotional state. Future studies should consider using hybrid methods and taking measurements before and after the emotion-inducing stimulus.

Another limitation is related to the time delay. Our methods align with previous laboratory studies that provided a short interval between encoding and retrieval (in this case, ca. 20 minutes), while findings from the neurobiology field suggest that at least a 24h interval between witnessing the event and memory retrieval is needed for long lasting memory formation (Sauerland et al., 2016). This is due to the involvement and role of stress hormones in long lasting memory formation (Sauerland et al., 2016). Furthermore, in a real-life situation the delay between the crime witnessed and the suspect identification procedure can vary from days to months (Flowe et al., 2018).

A number of studies have shown the importance of controlling this variable for recognition tasks (Snow, 2022; Lin et at., 2019; Sharot & Yonelinas, 2008; Park, 2005). Lin and collaborators (2019) studied, among other variables, the effect of time delay between witnessing a crime and the facial recognition task. Results showed that participants were less likely to identify the suspect after a 3-day interval. Adding the emotional component, Snow (2022) found similar results, where participants (in total, regardless of being in negative or neutral condition) were less likely (14.8%) to correctly identifying the suspect after a 1-week delay than after an immediate recognition (53.2%).

In a meta-analysis (48 studies, n = 3143), Park (2005) found that lower arousal contributed to better memory performance only after a short delay (two minutes) while higher arousal led to better memory performance for delays of 20-45 minutes, or 24 hours or more. Similarly, Sharot and Yonelinas (2008) suggested that emotional content enhances

recognition performance after a 24h delay when compared to neutral content. Hence future studies should consider adding a longer delay between study and test to better evaluate the impact of negative emotion.

Another important aspect worth mentioning is that the results observed in the current study should only be applied to types of crimes that do not present higher involvement of the eyewitness (or in which they were not victims themselves). This aspect was reflected in the intensity of emotional ratings of our participants, which were close to the medium levels of the scale. This could reflect a distancing subjects felt from the video content. A higher degree of involvement of the participants (e.g., a higher perceived treat for their own life) would allow for the examination of this variable in future studies.

Regarding the interpretation of results from this thesis, readers should be careful when generalizing the data. Besides the limitations presented before, attention should be paid to the definition of negative emotion when comparing different studies, as certain terms have been used as synonyms but can have different meanings (Głomb, 2022), as well as to which emotions were used (have they been specified by the authors, or they just analyze general distress?). Additionally, it is important to notice that current method provided a short interval between encoding and retrieval (ca. 20 minutes), which does not reflect usual delays observed in real life.

As shown before, the effect of negative emotion on memory is mixed and the application of the term "negative emotion" is vague. First, "negative" could be used to mean specific discrete emotions or in relation to the pleasant/unpleasant spectrum – and both models imply different physiological reactions, purposes, and behaviors. Secondly, the term "emotion" is commonly interchanged with "mood", "stress" and "arousal": but those words can have different meanings. Thus, such divergences lead to a high heterogeneity of studies being categorized together.

Accordingly, future investigations should be specific about what is considered as negative, if the origin of the emotional content was related to the task or stimuli (e.g., was it induced through mock videos to be analyzed later or through unrelated mathematical questions?), what aspects are being considered for memory (central/peripheral details? Both? What is the definition of such terms?), and which aspect of memory is being assessed (recall or recognition?). Such measures will contribute to a better understanding of the effect of negative emotion on memory and facilitate the comparison of studies in the literature, by making sure the researches evaluated the same components.

Conclusion

In addition to studying and contributing to the growing number of findings pointing to the (null) effect of negative emotion on eyewitness memory, and more specifically on facial recognition, this thesis shed light on an optional method for measuring suspect recognition. The PAIR comparison procedure seems to have a great potential to bring more reliable results for the justice system, as participants showed much higher accuracy in a PAIR lineup compared to the traditional SIMULTANEOUS lineup, irrespective of the emotionality of the witnessed event. Our results therefore point to a need to consider methodological advances in testing eyewitness memory for improving facial recognition.

References

Akan, M., Robinson, M. M., Mickes, L., Wixted, J. T., & Benjamin, A. S. (2021). The effect of lineup size on eyewitness identification. *Journal of Experimental Psychology: Applied, 27*(2), 369–392. https://doi.org/10.1037/xap0000340

Brewer, N., & Doyle, J. (2021). Changing the face of police lineups: Delivering more information from witnesses. *Journal of Applied Research in Memory and Cognition, 10*(2), 180–195. https://doi.org/10.1016/j.jarmac.2020.12.004

Brewer, N., & Wells, G. L. (2009). Obtaining and Interpreting Eyewitness Identification Test Evidence: The Influence of Police-Witness Interactions. In R. Bull, T. Valentine, & T. Williamson (Eds.), *Handbook of Psychology of Investigative Interviewing* (pp. 205–220). Wiley-Blackwell. https://doi.org/10.1002/9780470747599.ch12

Campos, T. F., Barroso, M. T. M., & de Lara Menezes, A. A. (2010). Encoding, storage and retrieval processes of the memory and the implications for motor practice in stroke patients. *NeuroRehabilitation, 26(2), 135–142.* https://doi.org/10.3233/NRE-2010-0545

Christianson, S.-Å., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition & Emotion*, *5*(2), 81–108. https://doi.org/10.1080/02699939108411027

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323–338. https://doi.org/10.1016/S0079-6123(07)00020-9

Davis, J. P., Maigut, A. C., Jolliffe, D., Gibson, S. J., & Solomon, C. J. (2015). Holistic Facial Composite Creation and Subsequent Video Line-up Eyewitness Identification Paradigm. *Journal of visualized experiments : JoVE*, (106), e53298. https://doi.org/10.3791/53298

Death Penalty Information Center. (2021). *DPIC Special Report: The Innocence Epidemic*. https://documents.deathpenaltyinfo.org/pdf/The-Innocence-Epidemic.pdf

Defensoria Pública do Estado do Rio de Janeiro. (2022). *Relatório consolidado sobre reconhecimento fotográfico em sede policial* [Consolidated report on photographic recognition at police headquarters].

https://www.defensoria.rj.def.br/uploads/arquivos/92d976d0d7b44b338a660ec06af008fa.pdf

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*(6), 687–706. https://doi.org/10.1007/s10979-004-0565-x

Dioso-Villa, R. (2015). A Repository of Wrongful Convictions in Australia: First Steps Toward Estimating Prevalence and Causal Contributing Factors. https://research-repository.griffith.edu.au/bitstream/handle/10072/123452/Dioso-VillaPUB1053.pdf?isAllowed=y&sequence=1

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & Cognition, 25*(2), 129–139. https://doi.org/10.3758/bf03201107

Douglass, A.B. & McQuiston-Surrett, D. (2006). Post-identification feedback: exploring the effects of sequential photospreads and eyewitnesses' awareness of the identification task. *Appl. Cognit. Psychol., 20*: 991–1007. https://doi.org/10.1002/acp.1253

Dutton, A., & Carroll, M. (2001). Eyewitness testimony: Effects of source of arousal on memory, source-monitoring, and metamemory judgments. *Australian Journal of Psychology*, *53*(2), 83–91. https://doi.org/10.1080/00049530108255128

Forgas, J. P., Laham, S. M., & Vargas, P. T. (2005). Mood effects on eyewitness memory: Affective influences on susceptibility to misinformation. *Journal of Experimental Social Psychology*, *41*(6), 574–588. https://doi.org/10.1016/j.jesp.2004.11.005

Flowe, H. D., Carline, A., & Karoğlu, N. (2018). *Testing the reflection assumption. The International Journal of Evidence & Proof, 22(3), 239–261.* doi:10.1177/1365712718782996

Gepshtein, S., Wang, Y., He, F., Diep, D., & Albright, T. D. (2020). A perceptual scaling approach to eyewitness identification. *Nature Communications, 11*(1), 3380. https://doi.org/10.1038/s41467-020-17194-5

Ghani, K. A., & Ning, L. C. (2016). Does Talking about Emotion Help Eyewitness Memory? The Role of Emotional and Factual Retelling in Memory Accuracy. *Journal of Cognitive Sciences and Human Development*, *1*(2), 1–10. https://doi.org/10.33736/jcshd.193.2016

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*(1), 1–67. https://doi.org/10.1037/0033-295X.91.1.1

Głomb, K. (2020). Does Eyewitness Guess or Recognize? Bootstrapping Face and Object Identification Accuracy. *Psychology and Law*, *10*(3), 73–85. https://doi.org/10.17759/psylaw.2020100306

Głomb, K. (2022). How to improve eyewitness testimony research: Theoretical and methodological concerns about experiments on the impact of emotions on memory performance. *Psychological Research, 86*(1), 1–11. https://doi.org/10.1007/s00426-021-01488-4

Harmon-Jones, E., Harmon-Jones, C., & Summerell, E. (2017). On the Importance of Both Dimensional and Discrete Models of Emotion. *Behavioral Sciences*, *7*(4), 66. https://doi.org/10.3390/bs7040066

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods, 54*(2), 556–573. https://doi.org/10.3758/s13428-019-01283-5

Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition*, *18*(5), 496–506. https://doi.org/10.3758/BF03198482

Houston, K. A., Clifford, B. R., Phillips, L. H., & Memon, A. (2013). The emotional eyewitness: The effects of emotion on specific aspects of eyewitness recall and recognition performance. *Emotion, 13*(1), 118–128. https://doi.org/10.1037/a0029220

Hulse, L. M., Allan, K., Memon, A., & Read, J. D. (2007). Emotional Arousal and Memory: A Test of the Poststimulus Processing Hypothesis. *The American Journal of Psychology*, *120*(1), 73. https://doi.org/10.2307/20445382

Innocence Project. (2020 April 15). *How Eyewitness Misidentification Can Send Innocent People to Prison*. https://innocenceproject.org/how-eyewitness-misidentification-can-send-innocent-people-to-prison/

Izard, C. E. (2010). The Many Meanings/Aspects of Emotion: Definitions, Functions, Activation, and Regulation. *Emotion Review, 2*(4), 363–370. https://doi.org/10.1177/1754073910374661

Kemp, R. I., Pike, G. E., & Brace, N. A. (2001). Video-based identification procedures: Combining best practice and practical requirements when designing identification systems. *Psychology, Public Policy, and Law, 7*(4), 802–807. https://doi.org/10.1037/1076-8971.7.4.802

Lenth, R (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.8.5) [Computer Software]. https://CRAN.R-project.org/package=emmeans

Lin, W., Strube, M. J., & Roediger, H. L., 3rd (2019). The effects of repeated lineups and delay on eyewitness identification. *Cognitive research: principles and implications*, *4*(1), 16. https://doi.org/10.1186/s41235-019-0168-1

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus." *Law and Human Behavior*, *11*(1), 55–62. https://doi.org/10.1007/BF01044839

Marr, C., Quaedflieg, C. W. E. M., Otgaar, H., Hope, L., & Sauerland, M. (2021). Facing stress: No effect of acute stress at encoding or retrieval on face recognition memory. *Acta Psychologica*, *219*, 103376. https://doi.org/10.1016/j.actpsy.2021.103376

Mittal, S., Singh, T., Arya, Y. K., & Tiwari, M. K. (2013). Effect of emotional arousal on eyewitness memory. *Indian Journal of Social Science Researches*, 10, 24 -30.

Mujawar, S., Patil, J., Chaudhari, B., & Saldanha, D. (2021). Memory: Neurobiological mechanisms and assessment. *Industrial psychiatry journal, 30*(Suppl 1), S311–S314. https://doi.org/10.4103/0972-6748.328839

Munsterberg, H. (1908). *On the Witness Stand*. McClure: New York.

Osborne, D., & Davies, P. G. (2012). Eyewitness identifications are affected by stereotypes about a suspect's level of perceived stereotypicality. *Group Processes & Intergroup Relations*, *16*(4), 488–504. https://doi.org/10.1177/1368430212454927

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior, 36(*3), 247–255. https://doi.org/10.1037/h0093923

Park, J. (2005). Effect of Arousal and Retention Delay on Memory: A Meta-Analysis. *Psychological Reports*, *97*(2), 339–355. https://doi.org/10.2466/pr0.97.2.339-355

R Core Team (2022). *R: A language and environment for statistical computing* (Version 4.2.2) [Computer Software]. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Red Inocente. (2023). *Nosotros* [Us]. https://redinocente.org/nosotros/

Saraiva, R. B., Souza, L., Nogueira, R., Coelho, L., & Alarcão, L. (2021). O Efeito de Carga Emocional e Intervalo de Retenção na Memória de Testemunhas Oculares [The Role of Arousal and Retention Interval on Eyewitness Memory]. *Psicologia: Teoria e Pesquisa*, *37*, e37211. https://doi.org/10.1590/0102.3772e37211

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 performance. *Behavioral Sciences & the Law, 34*(4), 580–594. https://doi.org/10.1002/bsl.2249

Sharot, T., & Yonelinas, A. P. (2008). Differential time-dependent effects of emotion on recollective experience and memory for contextual information. *Cognition*, *106*(1), 538–547. https://doi.org/10.1016/j.cognition.2007.03.002

Singmann, H., Bolker, B., Westfall, J., Aust, F. & Ben-Shachar, M. S. (2023). *afex: Analysis of Factorial Experiments* (Version 1.2-1) [Computer Software]. https://CRAN.R-project.org/package=afex

Snow, M. D. (2022). *Negative Emotion and Eyewitness Memory*. [Doctoral dissertation, University of Ontario Institute of Technology]. Ontario Tech Library. https://ir.library.ontariotechu.ca/handle/10155/1591

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: a meta-analytic comparison. *Law and human behavior, 25*(5), 459–473. https://doi.org/10.1023/a:1012888715007

Stridbeck, U., & Magnussen, S. (2012). Prevention of wrongful convictions: Norwegian legal safeguards and the criminal cases review commission. *University of Cincinnati Law Review, 80*, 1373–1390.

Taiwanese Innocence Project. (2022 August 27). *Zhǐrèn zhèngjù zhī zhìdù gǎigé* [System Reform of Identification Evidence]. https://twinnocenceproject.org/tipac2022_note04/

Talarico, J. M., Berntsen, D., & Rubin, D. C. (2009). Positive emotions enhance recall of peripheral details. *Cognition & Emotion*, *23*(2), 380–398. https://doi.org/10.1080/02699930801993999

Technical Working Group for Eyewitness Evidence (1999). *Eyewitness evidence: a guide for law enforcement*. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs. https://www.ojp.gov/library/publications/eyewitness-evidence-guide-law-enforcement

The Innocence Network. (2021). *Network Member Organization Locator and Directory*. https://innocencenetwork.org/directory

The National Registry of Exonerations. (2022). *2021 Annual Report*. https://www.law.umich.edu/special/exoneration/Documents/NRE%20Annual%20Report%202021.pdf

Trawiński, T., Aslanian, A., & Cheung, O. S. (2021). The effect of implicit racial bias on recognition of other-race faces. *Cognitive Research: Principles and Implications*, *6*(1), 67. https://doi.org/10.1186/s41235-021-00337-7

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis (*Version 3.4.1) [Computer Software]. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H. & Seidel, D. (2022). Scales: Scale Functions for Visualization (Version 1.2.1) [Computer Software]. https://CRAN.R-project.org/package=scales

Wong, H. K., Stephen, I. D., & Keeble, D. R. T. (2020). The Own-Race Bias for Face Recognition in a Multiracial Society. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/article/10.3389/fpsyg.2020.00208

Yaros, J. L., Salama, D. A., Delisle, D., Larson, M. S., Miranda, B. A., & Yassa, M. A. (2019). A Memory Computational Basis for the Other-Race Effect. *Scientific Reports*, *9*(1), 19399. https://doi.org/10.1038/s41598-019-55350-0