BRAIN SCANS IN THE COURTROOM: ANALYZING THE GENERAL PERCEPTION OF FMRI MEMORY RECONSTRUCTION EVIDENCE

by

EMILY CARLSON

A THESIS

Presented to the Department of General Science
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2020

An Abstract of the Thesis of

Emily Carlson for the degree of Bachelor of Science
in the Department of General Science to be taken June 2020

Title: Brain Scans in the Courtroom: Analyzing the General Perception of fMRI Memory Reconstruction Evidence

Approved: *Dr. Nicole Dudukovic*
Primary Thesis Advisor

This thesis seeks to assess the public perception of a new functional magnetic resonance imaging (fMRI) facial memory reconstruction technology and its potential future usage in creating facial composites that could serve as criminal evidence. To accomplish this, we presented participants with a facial composite posed as evidence in a hypothetical criminal court case. Participants read one of two accompanying descriptions of how that facial composite was developed: (1) with fMRI memory reconstruction technology that scanned the brain of an eyewitness, or (2) with modern computer software where the eyewitness built the facial composite from memory. In order to measure and compare the participants' perception of the two facial composite techniques, we presented them with a picture of the suspect's face along with the facial composite and asked them to rate the suspect's guilt, their confidence that the facial composite was actually the suspect, and the degree of similarity between the facial composite and the suspect's face. Though ratings for each question were higher in the fMRI condition compared with the computer software condition, implying a possible increased level of trust, none of the differences were statistically significant. These results, therefore, serve as good news for this developing fMRI technology, such that its potential future usage in developing facial composites may not be subject to any harmful biases.

Acknowledgements

Firstly, I would like to thank my primary advisor, Dr. Nicole Dudukovic. Thank you for being so patient and kind throughout this whole process, and for equipping me with all the tools I needed every step of the way. It has such been a privilege to work with someone as brilliant as you and I am so grateful for it. Thank you to Dr. Brice Kuhl for graciously agreeing to be on my thesis committee and for doing all this incredible fMRI research in the first place! Your work is inspiring, and it has been a great experience being a part of the Kuhl Lab as a research assistant. Thank you to Neeka Bayat-Barooni for being such a terrific Human Physiology GE and for so kindly agreeing to be on my committee as my third reader. I appreciate you so much! Thank you to Alex Tremblay-McGaw for your patience and willingness to help; it is so appreciated, and I would not have been able to collect all this data without you! Thank you also to the research assistants in the Kuhl Lab who all so kindly helped me collect data. There is no way we could have gotten this sample size without all of your help and I appreciate it so much! Thank you to Dr. David Frank for pushing me outside my comfort zone early on in this thesis process. You allowed me to think about my thesis in a whole new way and a build a solid foundation for it.

Thank you to Shadee Aghel for always being my study buddy—from Gen Chem to theses. We have shared a lot of tears from stress and late nights in the Knight Library, but we have shared way more tears from laughing at our absolutely hilarious jokes. I could not have survived college without you!! Thank you to all my roommates for being the most incredible and inspiring group of ladies and for always motivating me to dream big. Thank you to my amazing boyfriend, Logan, for supporting me unconditionally, for always pushing me to be the best version of myself, and for reminding me of what’s important in life—I appreciate you endlessly. And lastly, thank you to my parents for supporting me my entire life and always making me feel loved. I am beyond grateful for having such a wonderful family.

Table of Contents

List of Figures

List of Tables

Introduction

Imagine that the year is 2025, and you get a letter in the mail summoning you for jury duty as Juror #45. When the day finally arrives, you get interviewed and assigned to a case dealing with an armed robbery at a local hardware store. While sitting in the courtroom, you feel consumed by the tense, high-stakes environment of the trial. You see the beads of sweat rolling down the defendant's face as they are clearly running through all the worst-case scenarios for themselves and envisioning their potential fate at a state prison. A few moments later, the prosecutor hands the jury a facial composite which looks a little bit like the defendant. While handing it over, the prosecutor explains that it was developed using new "state-of-the-art technology" that allows neuroscientists to reconstruct eyewitness facial memories based off of the results of a functional magnetic resonance imaging (fMRI) brain scan. Given this piece of evidence, how would you feel about it? Would you feel compelled to trust it, knowing it was developed directly from the eyewitness's brain activity and memory of the culprit's face?

The technology used to create that facial composite is still in the developing stages (Lee & Kuhl, 2016) and is certainly not yet ready to be used in a court of law. Assessing the ethical implications of its potential future usage, however, serves as a crucial aspect of its development process. Studies have already displayed the ethical pitfalls of using facial composites in the context of eyewitness testimony (Charman, Gregory, & Carlucci, 2009), and many have also looked at the compelling nature of neuroscientific jargon (Weisberg, Keil, Rawson, & Gray, 2008; Weisberg, Taylor, & Hopkins, 2015; Fernandez-Duque, Evans, Christian, & Hodges, 2015). Our study, however, assesses the perception and potential consequences of using fMRI to reconstruct facial memories, as modern memory reconstruction techniques will likely become increasingly prevalent in the courtroom.

In this current study, we placed participants in the position of someone analyzing criminal evidence. From here, we tested for any persuasive effect that evidence may carry as a result of being derived from neuroscientific origins. By understanding public perception of fMRI evidence, both neuroscientists and those in a legal setting may better understand the ethical implications of bringing brain imaging technologies into the courtroom. Down the road, this could assist in generating guidelines on how this type of evidence should be used and presented in a trial, or if it should be used at all.

The present study is not attempting to display the strengths of using fMRI to reconstruct facial memories, although this technology may have the potential to produce quality facial composites that could potentially bypass various issues with current methods. Rather, the focus of the current research is to explore the perceptions of this technology and the potential implications of those perceptions, particularly in a courtroom setting. That way, when you are sitting in the juror’s seat and get handed that facial composite, it is not a one-way ticket to jail for the defendant, but rather another piece of quality evidence that must be critically analyzed.

Background

Eyewitness testimony has traditionally been fraught with problems. According to a 2007 review by Wells and Hasel, facial composites can harm innocent suspects who may coincidentally look similar to the facial composite, thereby potentially resulting in false convictions. Additionally, eyewitnesses tend to identify suspects who resemble the composite that *they* created, rather than the original face they saw. This finding has vast implications, such that people may trust their often-faulty memory over reality. The way people perceive and trust evidence greatly influences trial outcomes. Consequently, we are interested in assessing if people trust the fMRI-based memory reconstruction technique of the current study more than a fictitious non-fMRI counterpart.

Additionally, a 2009 study by Charman et al. suggests that a preconceived belief in a suspect's guilt can lead mock-jurors to interpret facial composites as highly similar to an innocent suspect, thereby heightening beliefs in the suspect's guilt. This may eventually lead to innocent people being wrongly convicted, perhaps also resulting in decreased efforts to identify the actual perpetrator. fMRI-based memory reconstructions could bypass this issue if precision of the facial composite was increased so greatly that connection to anyone besides the actual perpetrator would be next to impossible. Conversely, however, and perhaps more likely, the fMRI-developed facial composite could inflate this issue if viewers of the image have unjustifiable trust in it merely due to its neuroscientific origins.

Wells and Hasel (2007) also report that facial composites created by more traditional means, in which eyewitnesses build the composite themselves on an individual-feature level or describe their facial memory to an artist, can lead to major inaccuracies in the resultant facial composite. This occurs as people likely understand and recall faces more holistically (rather than by individual features), thereby proving it difficult to build accurate facial composites when done so in pieces. Modern computer software programs that address this issue by building facial composites through a more whole-face approach do exist, however, and are utilized by police forces. EvoFIT©, for instance, builds faces as eyewitnesses first look at a pool of randomly generated faces, then choose the face that most closely resembles the face that they remember. This chosen face becomes the parent to a newly generated pool of mutations of the parent face. This process continues until the eyewitness cannot choose faces any further as they all equally resemble their memory ("How it Works?", 2019). In 2010, EvoFIT was used by Humberside police and led to a 60% suspect identification rate (21 arrests from 35 composites)—quadruple that of the previously used feature-based system, E-FIT (Frowd et al., 2012). Though this is evidently an improvement to traditional facial composite techniques, developing fMRI technology by Lee and Kuhl (2016) could go a step further by skipping the eyewitness description and artist completely, and instead only utilize the eyewitness's brain activity.

These fMRI-based reconstruction methods have been developing in stages. In 2008, Kay, Naselaris, Prenger, and Gallant used a decoding technique defining the relationship between visual stimuli and fMRI activity to reconstruct an image from an observer's visual experience and identify which picture was viewed from a set of completely novel images. These findings helped set the stage for memory reconstruction based on fMRI activity alone. Cowen, Chun, and Kuhl (2014), however, first assessed if brain activity patterns were sufficient for reconstructing perceived facial images. This was done by utilizing a machine learning algorithm to map identified facial components with fMRI activity collected from the viewing of different faces. Subsequent fMRI activity was used to predict component scores; these scores then got translated to reconstructed images. Lee and Kuhl (2016) then further developed the technology in an attempt to reconstruct facial memories, as opposed to perceived faces. Their study demonstrates that facial perceptions can be reliably reconstructed from the activity pattern in the occipitotemporal cortex as well as several lateral parietal subregions. This was accomplished by applying a principal component analysis to a set of face images in order to create eigenfaces. These eigenfaces were then associated with various fMRI activity patterns such that predicted eigenface values could be used to reconstruct facial images. The current study seeks to assess the public perception of this neuroscience-derived fMRI facial image reconstruction technology, particularly in the context of utilizing it to produce facial composite evidence in criminal cases.

Just like most things in the world, neuroscience can easily be oversimplified and over-trusted. Consequently, the ethical implications of its general perception must be critically analyzed. Weisberg et al. (2008) demonstrated that non-expert groups perceive explanations containing neuroscience as more satisfying than explanations without it, even when the neuroscience information is irrelevant. Additionally, they found that in faulty explanations, the presence of neuroscience information can often mask the presence of otherwise obvious flaws in the explanation. Rhodes, Rodriquez, and Shah (2014) also found that when participants were reading news articles about scientific studies, the presence of neuroscience information resulted in higher ratings of both the study quality and the article quality.

Beyond just reading explanations, Ali, Lifshitz, and Raz (2014) assessed the overall trustworthiness of neuroscientific technology as they placed students in a mock neuroimaging device that supposedly decoded the participant's resting brain activity and read their mind. The study found that the majority of participants believed the entire paradigm, and that students who were enrolled in a psychology class focused on the shortcomings of neuroscience were only slightly more skeptical. This study points to the dangerous nature of neuroscience—where even those familiar with the field may be swayed by its allure and supposed abilities.

In 2011, McCabe, Castel, Rhodes, and Ewing found that when potential jurors read vignettes of criminal cases where the criminal lies about committing the crime, the vignettes which contained fMRI lie detection evidence (rather than polygraph or thermal facial imaging lie detection evidence, or no lie detection evidence at all) lead to more guilty verdicts. Gurley and Marcus (2008) also found that for not guilty by reason of insanity defenses (NGRI), the presence of neuroimages and brain lesions led to fewer guilty verdicts than those without neurological evidence, implying that the presence of brain-related evidence holds weight in a psychological testimony. These results suggest that jurors may trust evidence derived from fMRI technology merely due to the appeal of its origins. We are interested in seeing if fMRI-based memory reconstruction techniques will yield similar results, such that people trust a facial composite developed from fMRI technology more than a facial composite developed from modern computer software.

Collectively, these findings suggest that evidence derived via neuroscientific methods may be perceived as higher quality and more satisfying than their non-neuroscientific counterparts. In the context of a trial, this could be very problematic. Hopkins et al. (2016), however, found that people are not swayed exclusively by neuroscience, but rather swayed by reductive reasoning in general, when it is placed in an explanation of some phenomena. So, perhaps an fMRI-developed facial composite would not actually be much more persuasive than an advanced facial composite derived from another modern technique.

Aside from neuroscience jargon, the mere presence of brain images may inflict similar, if not stronger, biases. In a 2015 study, Weisberg and colleagues expanded on the seductive appeal of neuroscience from the Weisberg et al. (2008) study. The presence of neuroscience jargon was found as unnecessary for the biased results—the mere presence of brain images was compelling enough. McCabe and Castel (2008) found a similar effect of brain images—articles containing them were rated as better written and containing better scientific reasoning than those without brain images. Additionally, the presence of brain images also increased ratings of agreement with the articles' conclusions. Given that fMRI scans produce brain images (though not used directly as the facial composite), these results are pertinent to our study.

Interestingly, however, Fernandez-Duque et al. (2014) found that the presence of superfluous neuroscientific jargon increased the perceived quality of a psychological explanation, but the presence of a brain image had negligible effects. Michael, Newman, Vuorre, Cumming, and Garry (2013) also found that the presence of brain images had little to no effect on the perceived credibility of accompanying information and Schweitzer, Saks, Murphy, Roskies, Sinnott-Armstrong, and Gaudet (2011) found that neuroimages caused no effect on jury verdicts within NGRI defenses. Similarly, Hook and Farah (2013) observed that fMRI images did add appeal to research descriptions, but not due to scientific reasoning. In fact, when compared to non-brain-image photographs, fMRI images only improved interest levels, rather than scientific reasoning, surprise, innovation, or worthiness of funding. These results, though somewhat conflicting, exemplify that the presence of neuroscientific information and brain images may have compelling effects. The line between how advanced neuroscience information or image needs to seem in order to inflict bias remains inconsistent and unclear. By assessing whether fMRI-based facial composites elicit a compelling effect, our research will help to define that boundary.

If our study does find that people place more trust in an fMRI-based facial composite, it will support prior findings that suggest a compelling effect of fMRI technology, brain images, or just neuroscientific information in general. Consequently, it would become even more clear how critical it is to use neuroscientific technology, such as fMRI-based memory reconstruction, with caution and to assess its general perception prior to use in a non-research setting.

Existing research implies that a seductive allure may revolve around neuroscience in some sense or another. Through our research, we hope to better understand whether those biases exist within the context of fMRI facial memory reconstruction technology. To do this, we presented participants with a hypothetical crime scenario and a corresponding facial composite labeled as being based on the eyewitness's memory. In order to accurately assess the perception of this technology at its current stage in development, the facial composite used was an actual facial reconstruction image produced by Lee and Kuhl (2016). The facial composite was accompanied by one of two possible descriptions of how it was developed: (1) through this newly developing fMRI technology, which reconstructs the eyewitness' memory of the culprit based on their brain activity patterns, or (2) with a modern computer software, which allows eyewitnesses to build their own facial composite based on their memories. After participants viewed the facial composites, we measured their reactions using a questionnaire, and through that, compared their "levels of trust" in the facial composites. We hypothesized that participants would be biased towards having more trust in the fMRI-developed facial composite. Our hypothesis was based on the study by McCabe et al. (2011), who found that potential jurors were swayed by lie detection evidence when it was based on fMRI images. That study suggests that references to fMRI technology may compel the public and add an extra layer of trust. We are interested in assessing whether or not the reference to fMRI technology with regard to this developing facial memory reconstruction method would sway the public—particularly jurors—to an unwarranted and problematic degree.

Methods

Participants

For the in-person version of our study, one-hundred and thirteen University of Oregon undergraduate students recruited from the Human Subjects Pool participated in the study. All of these participants partook in a separate, existing study within the Kuhl Lab for credit and completed our in-person survey afterwards. For the online version of our study, three-hundred and nine University of Oregon undergraduate students enrolled in one of two introductory psychology courses completed the survey for credit online via Qualtrics. Online participants were between the ages of 18 and 44 years (average age = 19.5) with 156 participants identifying as female, 76 as male, and 2 participants identifying as non-binary or self-described. In terms of ethnicity, 156 participants identified as white, 34 as Asian, 25 as Hispanic origin, 7 as mixed or other ethnicity or origin, 6 as African American, 2 as Middle Eastern or North African, 2 preferred not to answer, 1 as American Indian or Alaska Native, and 1 as Native Hawaiian or other Pacific Islander. 75 participants in the online Qualtrics study did not report any demographic information due to an error by the survey administrator. One participant for the Qualtrics study was not included in the data as they provided a rating outside the given range.

Materials

For both versions of the study, we created the following hypothetical crime scenario:

> *A few weeks ago, an armed robbery took place at the cash register of a small, family-owned hardware store. The cashier of the shop witnessed the crime and was able to get a good look at the perpetrator. The robber ran away soon after the crime, and the cashier immediately called the police. From this cashier's memory of the robber, a facial composite was developed. Since then, a suspect has been identified and is now being held at the police station.*

We selected the "facial composite" image, as well as the "suspect" image from Lee and Kuhl (2016). The facial composite is an actual facial memory reconstruction produced by the researchers using this technique, and the suspect image is a photograph of the corresponding face.

We wrote two descriptions of how the facial composite was developed, whether through fMRI (condition "F") or a fictitious computer software named "fCIS" (condition "C"). The two descriptions had the same word count to avoid subjects being persuaded by length. Additionally, both of the descriptions *only* differed in the part that states whether the facial composite was developed via fMRI or fCIS in order to avoid subjects being swayed by other scientific jargon or explanations that are not consistent across both descriptions. The two descriptions read as follows:

Description for the fMRI "F" Condition:

The above facial composite was developed using a new state-of-the-art facial composite technique. With this technology, neuroscientists are able to reconstruct eyewitness' memories of a perpetrator's face based on the results of a brain scanning technique called Functional Magnetic Resonance Imaging (fMRI). While in the fMRI scanner, the cashier was instructed to make consecutive attempts to recall their memory of the *perpetrator's face, and a computer algorithm selected a set of faces from a large database that most closely matched the cashier's brain activity pattern. This resultant collection of faces resembling the perpetrator was then morphed into a single face to compute and build the above facial composite.*

Description for the Computer "C" Condition:

The above facial composite was developed using a new state-of-the-art facial composite technique. With this technology, eyewitnesses are able to reconstruct their memories of a perpetrator's face based on the results of a computer software program called Face Composite Imaging System (fCIS). While using the fCIS program, the cashier was instructed to make consecutive attempts to recall their memory of the perpetrator's face, and the cashier selected a set of faces from a large database that most closely matched their memory of the perpetrator's face. This resultant collection of faces resembling the perpetrator was then morphed into a single face to compute and build the above facial composite.

For the fMRI condition, the description was a simplified explanation of how this real technique works, lacking any of the technical details. For the computer condition, the description was of a fictitious software, reflecting the use of actual modern computer software which allows eyewitnesses to build their own facial composites from a bank of facial features ("How it Works?", 2019). Primarily, we aimed to have the two conditions differ in whether the facial image was developed directly from the eyewitness' brain activity, or if it was developed by the eyewitness recalling the face from their memory of the crime. This allowed us to test whether the mere fact that the image was developed from brain activity and brain imaging is inherently compelling (where the computer condition serves as the control), and thus gain a better understanding of the compelling nature of neuroscience.

Procedure

Participants were randomly assigned to either the fMRI condition, or to the computer software (control) condition. All participants first read the on-screen hypothetical crime scenario. Immediately after, the participants were prompted to look at the facial composite image that was displayed on the computer screen. The participant then read the description of how that facial composite was developed; half of the subjects read the description of the fMRI condition, and the other half read the description of the computer condition. For both conditions, the description was displayed on the computer screen. Next, participants were prompted to again look at the computer screen, but this time it was a side-by-side of the suspect image (a real photograph labeled as “Suspect”) as well as the same facial composite that they saw just prior. Participants then read a series of questions and responded by choosing a number on a 7-point Likert scale (1 representing the most negative response, and 7 representing the most positive response).

Initially, when we were piloting this experiment, we had only two questions for the participants: one that assessed their perception of how similar the facial composite seemed to the suspect’s face (1=extremely dissimilar, 7=extremely similar) and one that assessed their confidence that the facial composite was actually of the suspect whose face was displayed on the computer screen (1=extremely unconfident, 7=extremely confident). Participants also had the opportunity to answer a more open-ended question that asked them if they noticed anything throughout the study that they wished to comment on.

For the initial, pilot round of data collection, the descriptions and questionnaire part of the study were on paper, while the two images were on the computer screen. The participants were also the ones to fill out the questionnaire sheet. After some issues were observed (participants did not seem to consider the description when answering the questions and instead just hyper-analyzed the side-by-side images of the suspect's face and the facial composite on the computer screen), we adjusted the final procedure of the in-person study and put it all on a Microsoft PowerPoint presentation which was displayed on a computer screen. We also added another question (which was posed as the first question) and switched the order of the other two questions. The new question assessed how likely the participant believes the suspect to be guilty of the crime (1=extremely unlikely, 7=extremely likely), similar to the data collected by McCabe et al. (2011). In both the in-person and the online study, participants first answered the guilt question, then the confidence question, and lastly the similarity question (see Appendix).

For the in-person study, instead of having the participant fill out the questionnaire themselves while looking at the computer screen, the researcher transcribed their responses onto a paper questionnaire recording sheet. Additionally, in order to remind participants of how the facial composite was developed, we labeled the facial composite as either "Computer software composite" or "Brain imaging composite." Lastly, instead of having the images separate from the questions, we had each question on a separate slide, each of which was displayed underneath the repeating side-by-side of the suspect image and facial composite. Figure 1 is an example of a slide for the fMRI condition, question #1:

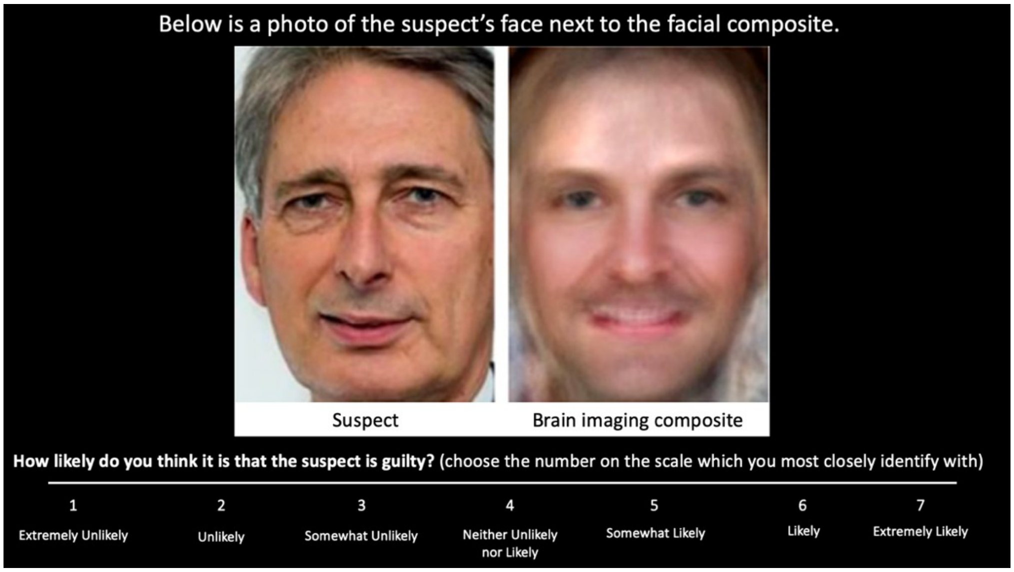

Figure 1: An example slide for the fMRI condition of the in-person study, question #1 (guilt question).

For all data collection taking place within the Kuhl Lab, results were collected via paper and pencil. Once we put the study into the general psychology survey, data were collected via Qualtrics. For this online study, participants were asked the same questions, and presented the same images (still labeled as "Suspect," "Brain imaging composite," and "Computer software composite"), but in order to remotely assess their engagement and acknowledgement of the independent variable, participants were also asked how the facial composite was developed. This question was constructed as a manipulation check and read as follows: "How was the facial composite developed? Please *briefly* describe below."

Results

Once data collection was completed, the data were analyzed using Microsoft Excel and SPSS. We computed the mean guilt, confidence, and similarity ratings for the two conditions, as well as their corresponding standard deviations. For each rating, we also calculated the proportion of participants who answered with a positive rating (5-7), negative rating (1-3), and an overall neutral rating (4). To determine whether there were differences between the two conditions for each rating, we ran unpaired equal variance *t*-tests. We assessed the interdependency of condition and rating (positive, negative, or neutral) with chi-square tests. Here, we only report analyses for all data combined (in-person and Qualtrics), but we also ran separate analyses for the data collected from the in-person study and the data collected from the Qualtrics study. The pattern of data and statistical results from those separate analyses were highly similar to the results from the combined data.

We first compared the mean ratings for the guilt question ("How likely do you think it is that the suspect is guilty?"), confidence question ("How confident are you that the person represented by the facial composite is the suspect?"), and similarity question ("How similar is the facial composite to the suspect's face?") between conditions. As shown in Table 1, for each of these three questions, the mean rating was higher in the fMRI condition than in the computer condition, as predicted by our hypothesis. However, none of these differences were statistically significant. Guilt ratings were numerically higher in the fMRI condition (M = 3.90, SD = 1.65) than in the computer condition (M = 3.74, SD = 1.56), but this failed to reach statistical significance, $t < 1$, $p = .331$. For the confidence ratings, the numbers were again higher in the fMRI condition ($M = 3.56$, $SD = 1.67$) than in the computer condition ($M = 3.42$, $SD = 1.50$), but again did not reach statistical significance, $t < 1$, $p = .360$. Lastly, the similarity ratings followed the same pattern, where ratings were not statistically significantly higher in the fMRI condition ($M = 3.64$, $SD = 1.74$) than in the computer condition ($M = 3.58$, $SD = 1.59$), $t < 1$, $p = .732$ (Table 1).

Table 1: Average ratings (out of 7) for each question compared between the computer and fMRI condition.

Condition	***N***	**Guilty Rating**	**Confidence Rating**	**Similarity Rating**
Computer	219	3.74	3.42	3.58
fMRI	203	3.90	3.56	3.64

Note that N differs across each condition. $N_{total} = 422$.

To assess the interdependency between condition and rating, we compared the proportion of participants for each condition who rated each question as negative (1-3), neutral (4), or positive (5-7). We then ran a chi-square test (Tables 2-4) but observed no statistical significance and therefore no dependency between condition and rating. For the guilt question, those in the fMRI condition were more likely to report that the suspect was guilty, as predicted by our hypothesis, but this failed to reach statistical significance, X^2 (2, $N = 422$) = 3.59, $p = 0.166$ (Table 2). Similarly, for the confidence question, those in the fMRI condition were more likely to report confidence in the facial composite being of the imaged suspect as we hypothesized, but this too failed to reach statistical significance, X^2 (2, $N = 422$) = 3.798, $p = 0.15$ (Table 3). Lastly, for the similarity question, those in the fMRI condition were again more likely to report positive values where they rated the facial composite as appearing similar to the suspect image as we hypothesized, but not to a degree that was statistically significant, X^2 (2, *N* = 422) = 4.042, p = 0.133 (Table 4). We also did a chi-square test for both the in-person and online studies separately, but the pattern remained close to the same and almost no statistical significance was found. The one exception was the chi-square test for the confidence question in the in-person study, which did reach statistical significance, X^2 (2, *N* = 113) = 8.360, p = 0.015. Participants in the fMRI condition were more likely to express confidence that the facial composite was the suspect than participants in the computer condition (.32 vs. .10 were the respective proportions). Because these results were so different than the results from the confidence question for the online Qualtrics study (p = 0.761), we suspect that this could be a statistical fluke rather than an actual indication of biases.

Table 2: Comparison of the proportion of negative, neutral, and positive responses to the guilty question.

	Not Guilty	**Neutral**	**Guilty**
Computer	.39	.28	.32
fMRI	.39	.21	.39

N_{total} = 422.

Table 3: Comparison of the proportion of negative, neutral, and positive responses to the confidence question.

	Not Confident	**Neutral**	**Confident**
Computer	.48	.27	.25
fMRI	.44	.23	.33

N_{total} = 422.

Table 4: Comparison of the proportion of negative, neutral, and positive responses to the similarity question.

	Not Similar	**Neutral**	**Similar**
Computer	.43	.27	.29
fMRI	.47	.19	.34

N_{total} = 422.

For the online Qualtrics study, we also examined responses to the manipulation check question ("How was the facial composite developed? Please *briefly* describe below."). We rated responses on a 3-point scale, where 1 = incorrect, 2 = partially correct, and 3 = correct. Responses scored a 1 if the participant completely ignored, misinterpreted, or incorrectly answered the question. Responses scored a 2 if they were on the right track but failed to state the precise method of facial composite development. To receive a 3, responses had to explicitly and accurately state the facial composite development method (e.g. "fMRI," "fCIS," "modern computer technology," "brain scanner," etc.), confirming that they were aware of their level of the independent variable. Interestingly, 64.6% of participants in the fMRI condition scored a 3, while only 48.1% in the computer (fCIS) group scored a 3. After sorting the responses based on score, we calculated mean guilt, confidence, and similarity ratings for each condition including only the responses that scored a 3 on the manipulation check. Similar to the pattern observed with the combined data, the mean rating for each condition increased from the computer group to the fMRI group, but still no statistical significance was found (*p*s > .1). We also ran chi-square tests for this collection of data, but again found a similar pattern to the combined data; for each question (guilt, confidence, and similarity), those in the fMRI condition were more likely to provide positive ratings. This trend, too, however, did not reach statistical significance (ps > .1).

Discussion

This study aimed to assess the public perception of a new functional magnetic resonance imaging (fMRI) facial memory reconstruction technology and its potential future usage in creating facial composites that could serve as criminal evidence. Results demonstrate that the average ratings were in the hypothesized direction where the average guilt rating was numerically higher when participants thought the facial composite came from an eyewitness's brain activity compared with when they thought the composite was generated using computer software. Similarly, participants were slightly more confident that the facial composite was the suspect and rated the two images as more similar when they thought the composite was generated using fMRI. This numerical pattern, particularly that of the guilt question, is consistent with the findings of McCabe, Castel, Rhodes, and Ewing (2011) who found that fMRI-based evidence (in their case, lie detection evidence) led to a greater number of guilty verdicts by mock-jurors than evidence created by non-fMRI techniques. However, the null hypothesis for our current study was ultimately supported as none of these differences were statistically significant. Consequently, these statistically insignificant findings may align more strongly with those of Schweitzer et al. (2011) who found that neuroimages caused no effect on jury verdicts with NGRI defenses.

It was also illustrated that for the guilt, confidence, and similarity question, a greater proportion of participants in the fMRI condition reported the suspect as guilty, were confident that the facial composite was of the suspect, and rated the facial composite as looking similar to the suspect image, respectively. These trends, however, again failed to reach statistical significance. Interestingly, results across both conditions illustrate that participants were generally skeptical of the facial composite; for all three questions, each of the average ratings fell below 4 (never surpassing a neutral average rating). This may be due to a limitation of the study design itself, however, as the pattern was consistent across conditions.

Though the results did not support our hypothesis, the lack of statistical significance and therefore absence of increasing trust in the fMRI-based facial composite may serve as good news for this new fMRI facial memory reconstruction technology. If the public is not actually compelled by this fMRI technology, then perhaps using it in a criminal trial as a means of creating evidence will be a safe procedure and not elicit biases, specifically within the jury. Though this fMRI facial memory reconstruction could bypass some of the issues with current facial composite techniques (i.e., bridge the gap between remembering a face and actually recreating that memory), it remains subject to the often-flawed eyewitness memory. Due to this, a healthy amount of skepticism is necessary in order to perceive this facial composite technique critically. Luckily, the generally low ratings given by participants to all three questions suggest that this skepticism may exist.

Despite the null hypothesis being supported, however, neuroscientists and those in a legal setting may still want to approach this technology with caution. Since the pattern of our data was in the hypothesized direction, it is possible that our sample size served as a limitation in our study and that with a larger sample size, the discrepancies would have reached statistical significance.

More likely, our manipulation may not have been strong enough. In particular, the image we used as the facial composite may have served as a limitation in our study.

Since we aimed to accurately portray the technology at its current stage in development, we used an actual facial memory reconstruction image from the Lee and Kuhl (2016) study. This facial image, however, may have appeared too unrealistic to elicit a compelling effect. It seemed that participants were often distracted by the differences between the facial composite and the suspect image rather than taking the independent variable (i.e., method of facial composite development) into account. This was evident in some of the answers to the manipulation check question for the online survey. For many of the responses which were scored as “incorrect,” participants commented on the quality of the facial composite as well as any differences they noticed between it and the suspect image rather than answering the actual question. Using a different image for the facial composite that more closely resembles the suspect image may lead to participants being more impacted by the description of how the facial composite was developed, and thus being more compelled by the fMRI-based technique.

In order to prevent experimental confounds, we also carefully controlled the description of how the facial composites were created with regards to both length and language. Though this was a necessary step to take in order to test the effect of only the independent variable, it may have inadvertently led to the fMRI-based description standing out less than it would in a more authentic courtroom setting. This is something to be cautious of when moving forward with research as the effect may be stronger in a study that more closely mimics a courtroom setting.

Additionally, differences between the in-person study and the online study may have skewed results. While participants in the in-person study were accompanied by a researcher who watched them complete the survey, participants in the online survey were alone and thus may have felt less pressure to thoroughly read the description of how the facial composite was developed as well as each of the questions. If this lack of incentive did exist, it may have skewed the data and led to ratings which demonstrate a less compelling effect. Even in the in-person study, however, the results failed to reach statistical significance thereby suggesting only negligible effects elicited by this possible lack of incentive.

If we were to design a follow-up to this study, we would likely use a larger sample size to see if this alters the results or brings the current pattern to statistical significance. Additionally, we could use a different image for the facial composite that more closely resembles the suspect image to see if this allows for the independent variable to more greatly affect participant responses. Lastly, we could convert the study to exclusively take place in-person in order to create a more consistent environment for participants and increase the incentive for them to complete the entire survey. If these adjustments were to lead to statistically significant results where participants were biased towards the fMRI condition, it would be interesting to complete another follow-up study where we question the validity of the fMRI technique. McCabe et al. (2011) found that when the validity of fMRI lie detection evidence was questioned, fewer guilty verdicts were given by the mock-jurors. This information would be valuable in the context of our study, particularly for developing guidelines on how this new fMRI facial memory reconstruction technique should be presented in a court of law, and if certain tactics like questioning its validity help to alleviate any biases.

Overall, the results of this current study support the null hypothesis and therefore indicate that in its current state, this new fMRI facial memory reconstruction technique does not need to spark alarm within the legal field in regard to any biases it may elicit when used to develop facial composites as criminal evidence. Though not statistically significant, the average ratings were consistently higher in the fMRI condition, thereby implying that researchers and those in a court of law should still approach this technology with some level of caution and strongly consider conducting further research on the topic.

Appendix

Below are the questions presented to participants in both the in-person and online studies:

How likely do you think it is that the suspect is guilty? (circle the number on the scale)

1	2	3	4	5	6	7
Extremely Unlikely	Unlikely	Somewhat Unlikely	Neither Unlikely nor Likely	Somewhat Likely	Likely	Extremely Likely

How confident are you that the person represented by the facial composite is the suspect? (circle the number on the scale)

1	2	3	4	5	6	7
Extremely Unconfident	Unconfident	Somewhat Unconfident	Neither Unconfident nor Confident	Somewhat Confident	Confident	Extremely Confident

How similar is the facial composite to the suspect's face? (circle the number on the scale)

1	2	3	4	5	6	7
Extremely Dissimilar	Dissimilar	Somewhat Dissimilar	Neither Dissimilar nor Similar	Somewhat Similar	Similar	Extremely Similar

Do you have any thoughts regarding this task that you would like to comment on?

Bibliography

Ali, S., Lifshitz, M., & Raz, A. (2014). Empirical neuroenchantment: From reading minds to thinking critically. *Frontiers in Human Neuroscience*, *8*(357), 1-4.

Charman, S., Gregory, A., & Carlucci, M. (2009). Exploring the Diagnostic Utility of Facial Composites: Beliefs of Guilt Can Bias Perceived Similarity Between Composite and Suspect. *Journal of Experimental Psychology: Applied*, *15*(1), 76-90.

Cowen, A., Chun, M., & Kuhl, B. (2014). Neural portraits of perception: Reconstructing face images from evoked brain activity. *Neuroimage*, 94, 12-22.

Fernandez-Duque, D., Evans, J., Christian, C., & Hodges, S. (2015). Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *Journal of Cognitive Neuroscience*, *27*(5), 926-944.

Frowd, C., Pitchford, M., Skelton, F., Petkovic, A., Prosser, C., & Coates, B. (2012). Catching Even More Offenders with EvoFIT Facial Composites. Paper presented at the Third International Conference on Emerging Security Technologies, Lisbon, Portugal.

Gurley, J. & Marcus, D., 2008. The effects of neuroimaging and brain injury on insanity defenses.(International Perspectives on Brain Imaging and the Law). *Behavioral Sciences & the Law*, *26*(1), pp.85–97.

Hook, C., & Farah, M. (2013). Look Again: Effects of Brain Images and Mind–Brain Dualism on Lay Evaluations of Research. Journal of Cognitive Neuroscience, *25*(9), 1397-1405.

Hopkins, E., Weisberg, D., & Taylor, J. (2016). The seductive allure is a reductive allure: People prefer scientific explanations that contain logically irrelevant reductive information. *Cognition*, *155*, 67-76.

How it Works? (2019). Retrieved from https://evofit.co.uk/how-it-works/

Kay, K., Naselaris, T., Prenger, R., & Gallant, J. (2008). Identifying Natural Images from Human Brain Activity. *Nature*, *452*(7185), 352-355.

Lee, H., & Kuhl, B. (2016). Reconstructing Perceived and Retrieved Faces from Activity Patterns in Lateral Parietal Cortex. *The Journal of Neuroscience*, *36*(22), 6069-6082.

McCabe, D., Castel, A., & Rhodes, M. (2011). The influence of fMRI lie detection evidence on juror decision-making.(functional magnetic resonance imaging). *Behavioral Sciences & the Law*, *29*(4), 566-577.

McCabe, D., & Castel, A. (2008). Seeing Is Believing: The Effect of Brain Images on Judgments of Scientific Reasoning. *Cognition*, *107*(1), 343-352.

Michael, R., Newman, E., Vuorre, M., Cumming, G., & Garry, M. (2013). On the (non)persuasive power of a brain image. *Psychonomic Bulletin & Review, 20*(4), 720-725.

Rhodes, R., Rodriguez, F., & Shah, P. (2014). Explaining the Alluring Influence of Neuroscience Information on Scientific Reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1432-1440.

Schweitzer, N.J. et al., 2011. NEUROIMAGES AS EVIDENCE IN A MENS REA DEFENSE: No Impact. *Psychology, Public Policy, and Law*, *17*(3), pp.357–393.

Weisberg, D., Keil, F., Goodstein, J., Rawson, E., & Gray, J. (2008). The seductive allure of neuroscience explanations. (Report). *Journal of Cognitive Neuroscience*, *20*(3), 470-477.

Weisberg, D., Taylor, J., & Hopkins, E. (2015). Deconstructing the seductive allure of neuroscience explanations. *Judgment and Decision Making*, *10*(5), 429.

Wells, G., & Hasel, L. (2007). Facial Composite Production by Eyewitnesses.(Author abstract). *Current Directions in Psychological Science*, *16*(1), 6-10.