5-10-2019

Finding prejudice in unexpected places: racially biased perception

Spencer Knafelc

FINDING PREJUDICE IN UNEXPECTED PLACES: RACIALLY BIASED PERCEPTION

by

SPENCER JAY KNAFELC

Under the Direction of Neil Van Leeuwen, PhD

ABSTRACT

A series of experiments reveal that people are more likely to mistake black men as holding a gun than white men (Payne 2001; Payne et al., 2002; Correll et al., 2002). These data suggest that real-world cases of racially biased object-identification occur, such as in cases of police killings of unarmed black men. The aim of this paper is to correctly model what goes on in people's heads, leading them to misidentify objects in these instances. One possibility is that people are making the wrong judgment about the object in question; perception might proceed as it should, but the viewer may *think* that they're seeing a gun due to a cognitive error. Instead, I present a model which construes the error as a result of a visual illusion: even though the object is a hand tool, erroneous visual processing causes them to have the illusory experience of a gun.

INDEX WORDS: Philosophy, Cognitive Science, Police Brutality, Perception, Implicit Bias

FINDING PREJUDICE IN UNEXPECTED PLACES: RACIALLY BIASED PERCEPTION

by

SPENCER JAY KNAFELC

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2019

FINDING PREJUDICE IN UNEXPECTED PLACES: RACIALLY BIASED PERCEPTION

by

SPENCER JAY KNAFELC

Committee Chair: Neil Van Leeuwen

Committee: Dan Weiskopf

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2019

ACKNOWLEDGEMENTS

Primarily, I wish to thank my advisor, Neil Van Leeuwen, for his feedback, criticism, encouragement, and contribution to this thesis. I am exceptionally grateful for his guidance. I am indebted to Dan Weiskopf and Eddy Nahmias for comments and feedback on previous drafts. I wish to thank my colleagues and friends that took part in our thesis workshop group: Sam Munroe, Chase Halsne, Emily Tilton, Hunter Cantrell, Nick Alonso, Andrew Gilley, and Nick Linhares. Finally, I owe thanks to the participants of our beloved writing group, in which the majority of my thesis was written, Rachel Lehmann, Tiffany Zhu, Nick Alonso, Emily Tilton, and our leader, Neil.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

In March 2018, police shot and killed Stephon Clark, a 22-year-old black man suspected of breaking into cars, while he was standing in the back yard of his grandmother's house. They fired twenty bullets, eight of which struck Clark. One officer claimed he had seen a gun in Clark's hand. Afterward, it was discovered that he did not have a gun but was instead holding a cell phone. Clark's case represents a larger pattern. Statistics reveal that police kill citizens of certain groups at higher rates than others. Black men are disproportionately represented as victims in the US (Mapping Police Violence, 2018). In many cases, as in Clark's, officers claim to have seen a weapon.[1]

Psychologists, furthermore, have experimentally demonstrated that people are more likely to mistake black men as holding a gun than white men in a series of different Weapon Identification Tasks. In a study by Correll et al. (2002), participants were tasked with "shooting" or "not shooting" individuals displayed on a screen in a computer simulation. The people depicted were either holding a gun or a harmless object, such as a cell phone. Participants were directed to "shoot" if the person depicted held a gun and to "not shoot" if the they were not. If the depicted person was a black man, participants were more likely to mistakenly "shoot" them. A similar, but more complex experiment run by Greenwald et al. (2003) demonstrated the same phenomenon. Participants were told to shoot citizens holding guns, to report safety if the individual encountered was a police officer, or to do nothing if the displayed individual was a citizen without a gun. Results revealed a bias against black men; whether they were police officers or citizens, participants were more likely to mistakenly "shoot" the depicted person if they were

[1] Philando Castile was shot to death in his car after being pulled over by an officer who claimed that he believed that Philando had grabbed his gun but was only grabbing his wallet. Saheed Vassell was shot and killed by police that took him to be holding a gun while he was holding a metal pipe.

black. Finally, Payne (2001) and Payne et al. (2005) presented participants with clear images of objects which were either guns or hand tools. They had under a second to report the object's identity as either a gun or non-gun but were first primed with a face that was either black or white. Participants were more likely to report that a hand tool was a gun if they were primed with a black face. These priming experiments will be a significant focus for this paper, and I will be referring to them as the Image Experiments to distinguish them from the other weapon identification tasks.

Collectively, the results of these experiments show that people have a racial bias that makes it more likely that they will think that an object is a gun if it's held by a black man. There is reason to think that the phenomenon observed in these experiments also occurs in real-world cases. Whenever police officers think that a citizen is holding a gun, it may be because of the same psychological mechanisms that generate racial bias in these experiments.[2]

There are several ways in which racial bias can manifest itself in the mind and so it is important to find out exactly how it arises in these cases of race-mediated object misidentification. Correctly modeling psychological mechanisms that cause racial bias will aid in developing strategies for correction, hopefully thereby preventing further atrocities such as the killing of Stephon Clark. Moreover, it can aid in normatively evaluating those that display such biases, and it may further our understanding of the structure of the mind more generally.

The aim of this paper is to correctly model what goes on in people's heads, leading them to misidentify objects in these instances. One possibility is that people are making the wrong judgment about the object in question; perception might proceed as it should, but the viewer may

[2] Of course, it is possible that police officers that claim that they had seen or believed that a suspected individual was holding a gun are lying. They may have known that it was not a gun. However, I am attempting to model those cases in which misidentification does occur and I presume that in many cases in which police officers claim that a gun is present, it is due to misidentification.

incorrectly think that they are viewing a gun due to a cognitive error. Another possibility is that they experience an illusion; due to a processing error, their visual system may lead them astray. In this case, despite the fact that the object is a hand tool, their visual system causes them to have the illusory experience of a gun.

Thus, we can start with the following questions: do officers and participants actually 'see'[3] a gun that's not there? What is perception's role in generating the error of object misidentification in the experiments that demonstrate the effects of racial bias? And what mental state causes racially biased behavior (like reporting the incorrect object, or pulling the trigger of a gun)? Answering these questions involves building a model of the cognitive architecture involved in generating mistaken reports and localizing the error to a specific process or set of processes. One way of categorizing the possible architectures is by examining whether the error generated is perceptual, post-perceptual, or the result of an interaction between perceptual and other cognitive processes.

In this paper, I consider three models that might explain object misidentification in weapon identification tasks. I spell out what I take to be the best versions of each model, along with their implications for cognitive architecture in general. First, I consider what I call the Executive Failure Model, the details of which have mostly been developed by Payne and colleagues (2005). It states that in the Image Experiments, perceptual processing outputs a correct representation of the object, but one's post-perceptual prejudice causes them to report that the hand tool is instead a gun. Then, I propose a model of my own, called the Illusory Percept Model, which says that one's bias results from perception misrepresenting the world. In other words, erring participants in these experiments actually have the experience of seeing a gun

[3] I will use 'see' throughout this paper to refer to the experience of perceivers where they have an experience as if they are viewing a gun despite their experience being in part caused by an object that is not a gun.

despite being presented with a different object. Finally, I assess the possibility that errors result from the generation of ambiguous percepts, where subjects, influenced by bias, make judgments about unclear perceptual feedback causing them to misidentify objects. I call this the Ambiguous Percept Model.

These three models are theoretically important candidate explanations for the phenomenon of racially biased weapon misidentification. In exploring this theoretical space, I elucidate the background cognitive architectures that each model requires. After spelling out each model, I consider which offers the best overall explanation of the relevant data. I argue that we ought to reject the Executive Failure Model on the grounds that it has bizarre consequences. I will argue that since the Illusory Percept Model adequately explains the available data and because it accords well with current theories of perception, such as Bayesian models like the Prediction Error Minimization (PEM) framework (Clark, 2013; Clark, 2015; Howhy, 2014), we have independent reason to favor it over the Executive Failure Model. I then provide reason to favor the Illusory Percept Model over the Ambiguous Percept Model by demonstrating the existence of what I call the Perceptual Resolution Principle. I conclude that the Illusory Percept Model is our current best option since it best explains the relevant data, avoids implausible implications, and coheres with current theories of perception.

My conclusion is significant for a few reasons. If weapon misidentification is caused by a post-perceptual error, then the issue could be solved by teaching people to have correct beliefs about the world; ridding of stereotypes that influence judgment will fix this particular manifestation of racial bias. However, if the issue is with perceptual processing, then it might not be so simple. The model that I defend suggests that observed bias can occur independent of the content of one's beliefs (implicit or explicit). Moreover, if the Illusory Percept Model is right,

questions arise regarding whether these illusions are inevitable for those whose visual systems are already disposed to generate such errors. I consider the implications of the Illusory Percept Model for training and preventing instances of such illusions. I also explore some of this epistemological and ethical issues raised by my conclusion.

2 EXECUTIVE FAILURE MODEL

In this section, I introduce what I call the Executive Failure Model. Roughly, this model states that there exist two separate processing streams – perceptual processing and a stereotyping process. One might suspect that if someone mistakenly believes that a black man is holding a weapon whenever they are instead holding a harmless object, something must be going wrong with their visual system – they're just seeing the world wrong. Instead, this model suggests that this isn't necessarily the case. Their visual system is working just fine, but a process in their mind that automatically and involuntarily stereotypes causes them to report that the object is a weapon due to bias that results from the stereotyping process.

Recall that in the Image Experiments, subjects are told that they are to identify an object as a gun or non-gun, and are then presented with the object. Given the simplicity of the task – being presented with a clear image and identifying it – it's shocking that errors are so prevalent. Nevertheless, participants are more likely to report that hand tools are guns if primed with a black face. The Executive Failure Model posits the presence of an automatic stereotyping process that's largely independent of perceptual processing to explain this phenomenon.

According to this model, for a participant to successfully report the identity of an object in any of these experiments will depend on whether she can rely on her perceptual system and exclude processes that generate racial stereotypes. One might picture two subprocesses in the head competing for influence over behavior. If one's behavior is controlled by their stereotyping

process rather than their perceptual system, it is due to a failure of executive control – the process responsible for planning and executing actions that are in line with one's goals (Payne, 2004). Executive control includes the ability to select for information that is relevant for completing the task at hand. In this case, allowing the stereotyping process to factor into one's behavior is not conducive to successful object identification. Thus, object misidentification results from a post-perceptual error where the executive function fails to stop stereotyping from playing a role in action, such as reporting the identity of an object. In what follows, I spell out the Executive Failure Model in more detail.

2.1 Stereotypes

Racial stereotypes[4] are an example of how an adaptive and valuable cognitive process can go awry, leading to detrimental effects. Many, if not most, instances of "stereotyping" are helpful and rational. However, whenever it comes to social cognition, they can be not only misleading and irrational, but they can have pernicious consequences at the individual and societal levels.

Consider the ease with which our cognitive systems identify objects in the world, subsuming them under categories and attributing features to them based on their apparent category membership. The latter tendency is the origin of stereotypes. Whenever we encounter a new object that shares apparent similarities with other objects that we know some things about, we often unconsciously assume that this new object shares some other features with those objects

[4] In this section, I refer often to a stereotype process as if stereotyping or bias can be associated with one particular mechanism. This is a gross idealization. Mechanisms responsible for overrating the size of black men compared to white men may be different than those responsible for observed bias in the Weapon Identification Task (Wilson et al., 2017). Notably, one's scores on the Implicit Association Task (Greenwald et al., 1998) did not correlate with one's performance on Payne's Weapon Identification Task (Mandelbaum 2014, 3). The mechanisms underlying implicit bias as it manifests in behavior may be disparate and complex. For example, out-group prejudices likely play a role in racial bias. It would be surprising if the mechanisms responsible for out-group bias are the same mechanisms responsible for specific stereotype judgments of ethnic groups. In the Executive Failure Model, the mechanism responsible for bias in the Weapon Identification Task is a distinct process from perception. Thus, whenever I refer to "distinct stereotyping process", it is only for the sake of convenience; I mean to refer to whatever collection of cognitive processes are responsible for producing bias.

as well. For instance, if we encounter a metal object of a particular shape with four tires, we will automatically take for granted that it contains an engine, is being operated by the person inside, that it requires gasoline, etc. If we see a small animal with four legs, beady eyes, and a bushy tail, we assume that it eats nuts and is capable of climbing trees. The process responsible for sorting the world in this way proceeds automatically and unconsciously. It allows creatures like us to make sense of the world by inferring features of available opportunities for action and it operates without controlled effort. Stereotyping so construed is clearly useful.

However, this process (or set of processes) is at least in part responsible for harmful stereotypes. Stereotypes that are applied to individuals based on their cultural and ethnic groups – whether they are true or false and regardless from where they originate – impact how one interacts with and view members of such groups (McConnell & Leibold, 2001; Cottrell & Neuberg, 2005). For example, people find black men more aggressive than white men (Sagar & Schofield, 1980). People that seem to harbor no conscious racist beliefs nevertheless display racial bias so long as they are aware of cultural stereotypes, an effect observed in Correll and colleagues' Weapon Identification Task (2002). Moreover, stereotyping can become warped by in-group/out-group biases. The way that many come to believe that Mexicans are criminals is not the same way that people come to believe that cats like fish. For instance, individuals perceive members of the out-group as more alike (having less variability among traits) than members of the in-group (Linville, Fischer, & Saolvey, 1989) and people are more likely to view their own ingroup as superior to the out-group (Brewer, 1979). These two biases enter into the information-processing systems that generate stereotypes, leading to fallacious and harmful generalizations.

The Executive Failure Model states that the system in the mind responsible for automatic stereotyping is distinct from perceptual processing. The automatic stereotyping process receives

information from perception, but then operates independently.[5] Behavior is either guided by perceptual processing without any effect from stereotyping, or it is guided by the automatic stereotyping process.

Let's apply this model to the Image Experiments. Participants were to identify an object presented on a screen as either a gun or non-gun (the objects were either guns or hand tools). The object to be identified was displayed for 200 milliseconds, and the participants had 500 milliseconds after the stimulus was removed to respond. They were primed with either a black or a white face, displayed for 200 milliseconds before the presentation of the stimulus. Participants were more likely to report that a hand tool was a gun if they were primed with a black face. The Executive Failure model states that participants who are presented with a hand tool but incorrectly report the object to be a gun are likely seeing the hand tool display just fine, but their automatic stereotyping process that associates black men with the concepts 'violence', 'crime', or something in this semantic proximity causes them to report that the object is a gun.[6] Automatic post-perceptual bias caused the participant to err.

2.2 Executive function

The model has given us an explanation for erroneous reports, but there must be something that explains successful 'non-gun' reports. When things go right, the Executive Failure model states that the reason is participants' successful exercise of executive control.

---

[5] One might think that the automatic stereotyping process is bypassing perception altogether. However, perceptual representation must be playing some causal role in behavior. To see this, consider if the stimulus presented were a bouquet of flowers rather than a hand tool. Would we see participants making mistakes at the same rate? The features shared between the hand tool and the gun must do some work in generating the misidentification of the hand tool as a gun. If the stereotyping process bypassed perception altogether, we would expect the same results even if *no* object were presented.

[6] It is possible – likely, I think - that cashing out the mechanism of racial prejudice merely in terms of unconscious conceptual associations is incorrect. Instead, it might be unconscious, propositionally structured belief-like states (Mandelbaum, 2014). This model can remain neutral regarding the specific post-perceptual mechanism that produces the conflict between accurate report and behavior-guided by stereotype.

Executive control is the ability to "plan and carry out selective behaviors in a way that follows one's goals" (Payne et al. 2005, 37). Part of planning and carrying out an action is selecting the information that will guide behavior; selecting that which is deemed most conducive to the task at hand. The emerging picture may be crudely put as a competition between the output of perceptual processing and the output of the automatic stereotyping process for guiding behavior. It is up to executive control to "selectively gate" the latter so that accurate visual representation can cause the subject to correctly report the hand tool as such.

Using this model, Payne and colleagues (2005; also, see Correll et al., 2002) have likened object misidentification in these experiments to erring subjects in the Stroop task. In the Stroop task, participants are presented with color words like 'blue' and 'yellow'. The words are printed in colored text that is usually incongruent with the word itself (e.g. the word 'green' printed in red ink). The participant must report the ink color but not the word. This proves to be difficult; it's difficult to stop oneself from blurting out the word rather than the ink color. It's the role of executive function to suppress the automatic impulse to report the word itself in order to report the ink color.

The idea is that there exists an analogy between what goes on in participants heads during the Stroop task and during the weapon identification task. It is the job of executive control to suppress task-irrelevant information – information processed and generated by automatic systems – that might interfere with the planning and carrying out of behavior. Just as one has trouble in the Stroop task because one automatically processes the word and thus must suppress reporting the wrong information, one has trouble in the object identification tasks because she must suppress stereotyping from impacting her report.

Thus, according to the Executive Failure Model, the reason some participants err in Weapon Identification Tasks is because they are unable to selectively gate their automatic stereotyping process from influencing their report, whereas those that succeed in using executive control report correctly. It provides us with a plausible picture of how mental operations might go wrong when trying to get the world right. In the next section, I'll present a competing model which locates error at the perceptual level.

3 ILLUSORY PERCEPT MODEL

Susanna Siegel, in *The Rationality of Perception*, examines the problem of what she calls "hijacked experience" (2016**).** An experiential state may be formed by irrational inference or might be influenced by other mental states such that its content is altered. If either of these events occur, causing the state to misrepresent the world, then the experiential state in question has been "hijacked": it's been influenced in a way that makes it epistemically problematic, just as beliefs that are generated from poor inferences are problematic. There are a number of plausible ways that perceptual experience can be "hijacked". For instance, if perception is inferential, the inferential process may utilize bad information or bad inferences, corrupting the content of resultant experience. Also, other mental states, such as moods, emotions, or belief-like states could interfere with perceptual states. We know that perception often goes awry, and there is reason to think that often, it's because conscious perceptual states have been "hijacked".

Consider the memory color effect, where one's "knowledge" of the color of familiar objects modulates the appearance of such objects (Hansen et al, 2006). In a set of experiments, participants were asked to adjust the color of familiar fruits until they appeared achromatic. Participants consistently adjusted the color such that it was shifted slightly too far in the opposite direction of the typical color as opposed to correctly stopping it on gray, suggesting that stored

information about the color of objects has an effect on one's perceptual experience. For instance, a banana was adjusted too far in the opposite direction of yellow, but a yellow object that isn't typically yellow would be adjusted so that it was entirely gray-scale.

Moreover, hallucinations can be conditioned into neurotypical subjects (as well as those with psychotic symptomology) (Powers et al., 2017). In these experiments, participants underwent classical conditioning where a tone was paired with a checkerboard visual stimulus. Eventually, the presence of the visual stimulus caused subjects to report hearing a tone. Because the participants' perceptual system had come to expect a tone accompanying the paired visual stimulus, subjects actually experienced hearing the tone. Both color memory and induced hallucinations show how stored information in one's cognitive system – whether they are belief-like states or stored perceptual expectations – can significantly alter how one experiences the world: such mental states can "hijack" experience.

Now imagine that you're a participant in one of Payne's experiments. You are asked to report whether an object presented on the screen is a gun or a non-gun, and as a result, you are anticipating the presence of a gun. You know that you have less than a second to report your answer. When an object is flashed before you, you literally have the experience of seeing a gun. This is what I suggest occurs. If this is true and we think that Weapon Identification experiments have ecological validity, then we should also expect this to be the case for some instances of fatal errors by police officers: instances where their biases and expectations cause them to see objects such as cell phones or metal pipes is if they were guns. In the following sections, I spell out some reasons to suspect that these visual hallucinations occur and mechanisms that might explain their occurrence.

3.1 Becoming biased: experience-altered perceptual assumptions

You might have misgivings about how exactly prejudice could manifest at the perceptual level without evidence that percepts vary between individuals due to experience-caused shifts in the operations of their perceptual systems. Here, I'll present evidence that environmental tuning of the visual system occurs, first by showing that information transmitted through other sense modalities seems to temporarily re-tune visual "assumptions", and then by providing evidence that sensitivity to visual illusions is prior experience-dependent in a way that varies between cultures.

It is accepted that the visual system uses stored information to make "assumptions" in order to disambiguate noisy input. One example of such an assumption is the light-from-above heuristic, encoded in the visual system to extract information about the shape of objects from shading. The idea is that the light source in a given environment tends to be the sun, and so the visual system takes this fact into account by "assuming" that light comes from above[7] to determine the features of objects in the world, such as whether a sphere is convex or concave. For our purposes, the question that arises is whether the "assumptions" of the visual system can be modified through experience.

In fact, it seems that the light-from-above heuristic can be altered by experience. Adams et al. (2004) show that one's light-from-above heuristic can be temporarily altered if given haptic feedback that conflicts with the initial assumption implicit in the visual system. Participants' pre-existing light "assumptions" were measured through judgments of "bump-dimple stimuli" where they were to report if circles, shaded to appear as if spheres, were convex or concave. Then, participants were trained with stimuli that were actually convex or concave where they were

[7] Approximately. See Sun & Perona, 1998.

allowed explore the stimuli with their fingers, receiving haptic feedback. Stimuli that looked as if they were concave due to shading were actually convex and vice-versa. The training stimuli altered the spheres to suggest that the light source has shifted either 30 degrees to the left or to the right. After training, participants were asked to judge the vision-only 2-d stimuli again. Their judgments about the circle's orientation reveal that participants implicitly "learned" that the light source had shifted in the manipulated direction. In other words, assumptions that factored into their perception of objects' shape were altered, influencing the content of participants' experience.

More evidence for experience-influenced perceptual "assumptions" comes from Bremner et al., (2016). The researchers compared the effects of the Ebbinghaus illusion (Figure 3.1) on children of the indigenous Himba children who live in rural Namibia, urban Namibian children, and children of the UK. The effects of the illusion were significant for UK children, whereas rural Namibian children showed a small effect. Urban Namibian children showed a greater effect than rural Namibian children, but less than UK children. The hypothesis is that cross-cultural differences are mediated by participants' exposure to urban environments.

Figure 3.1 The Ebbinghaus illusion.
The orange circles in the center are the same size, despite their appearance.

Regardless of the specific mechanism, the alteration of the light-from-above heuristic and the difference in the effect of the Ebbinghaus illusion explained above show that the visual

system's "assumptions" are altered by experience, tuning them to accommodate for the perceivers environment. The model I've put forward, the Illusory Percept Model, is another instance of a culturally conditioned illusion. Put simply, perceptual systems make "assumptions" and such assumptions result from the prior experiences of the perceiver. While there may be an important difference between "low-level" assumptions such as where light sources tend to be and "higher-level" assumptions like the identity of objects given specific social features, my present aim is only to provide an existence proof of malleable visual "assumptions" to motivate the hypothesis that such "high-level" assumptions exist. There is therefore support for the claim that one's perceptual "assumptions" can be tuned through experience in the fashion required for my proposed model to make sense. It remains an open question as to what degree and at how "high" of a level such attunement occurs.

Thus, we might think that there are ways in which exposure to certain events might cause one's perceptual system to become biased by having its assumptions altered based on something like skin color. For instance, it's plausible that exposure to certain types of media that depict black men wielding weapons could explain how perceptual "assumptions" become racially biased. Importantly, it should be noted that the mechanism through which perceptual "assumptions" become biased need not be cognitive penetration.

Here are two ways that perceptual processing can become biased to produce ethnicity-mediated object misidentification.

i) One lives one's life unreflectively consuming media that depicts black men as violent and soaking up ambient cultural stereotypes that black men are criminals, thereby forming beliefs, explicit or implicit, that Black men are more likely to be dangerous,

criminal, and are thus more likely to be wielding a weapon. These beliefs somehow interact with perceptual processing, influencing perceptual experience.

ii) Through exposure to media that depicts black men as violent, criminal, and more likely than white men to be wielding a weapon, one's perceptual system becomes attuned to such features and regularities, thereby influencing perceptual "assumptions". One's conscious beliefs about such matters has no impact on perception's operations.

The first scenario is an instance of the cognitive penetration of perceptual experience, meaning that higher-order belief-like states influence the operations of perceptual processing and therefore one's experience. The second scenario depicts the biasing of perceptual processing without the occurrence of cognitive penetration. Both of these scenarios, I suggest, are plausible. It is a matter of empirical investigation as to whether it's the case that the first, second, both, or neither occur. My point is that environmental tuning of perception can operate solely at the perceptual level – it doesn't have to be higher-order states affecting perceptual processing. (To be clear, I'm not ruling out cognitive penetration. The model that I'm putting forward can remain neutral as to whether it occurs. Having said that, this draws attention to the further possibility that the cognitive penetration occurs and exacerbates the effects of racially biased perceptual processing, further increasing the possibility of illusory experience.)

In the next section, I'll introduce what I call the Perceptual Resolution Principle, which motivates and constrains the theory that I am putting forth. I'll show that there are independent reasons to accept the principle's existence and how it fits with the Illusory Percept Model.

3.2 The Perceptual Resolution Principle

The suggestion that perception would go wrong enough to produce a representation of an entirely distinct object is striking. Participants in the Image Experiments are presented with a socket wrench, for example. Why would we think that perception would instead output a gun – an entirely different object? If there isn't enough detail or time to process the socket wrench, wouldn't the output just be a bit muddled or less detailed? In this section, I briefly argue for the existence of what I call the Perceptual Resolution Principle, which states that perceptual processing aggressively seeks to disambiguate incoming information in order to output determinate percepts rather than percepts with ambiguous content.

There are empirical and theoretical reasons to think that perceptual processing outputs almost entirely determinate percepts. The principle gains empirical support from the phenomenon of binocular rivalry, and theoretical support from considerations about the function of perception.

Imagine that you are looking at an image of a house and face, split so that each eye can see only one of the objects. What will your perceptual experience be like? It could be a jumbled mess, the result of the visual system's confusion due to an unfamiliar state of affairs. You might see a relatively coherent mash-up of house and face. Or, you might see only one image at a time; either a house or a face. The latter is what occurs, and the phenomenon is called binocular rivalry. The perceptual representation of the strange stimulus is always a determinate object. It's as if perception attempts to select the most plausible representation given the information it has; perceptual processing selects either house or face in favor of conscious perception making sense (Howhy, 2013 23-25).

The phenomenon of binocular rivalry reveals the imperative that guides the visual system: generate determinate representations despite noisy or unfamiliar sensory input. Input always involves noise and ambiguity. It would not be adaptive for creatures to have perceptual systems that output ambiguous or indeterminate representations since this would only confuse the organism. One is better off if their visual system constructs its best guess given the current available information, so that the world is organized into determinate and actionable entities. Avoiding indeterminate representations will come at a cost, however. Often, the best guess about a given state-of-affairs in the world will be incorrect, the result being misperception. The evolutionary trade-off is one between having actionable representations at the cost of getting the world wrong every once in a while.

3.3 Inferential perception

Theorists have long suggested that perception occurs as a process of unconscious inference (Helmholtz, 1867; Hatfield, 2002; Clark, 2015; Corlett, 2018). Rather than perception being a passive process of taking in sensory information, it is actively inferring the cause of proximal stimulation. From prior experience, the brain stores models of the world. These models are used by the perceptual system to make sense of incoming sensory signal. In this sense, perception is thought to be a process of unconscious inference – given models of the world based on prior experience, what is the most likely the cause of current sensory input? The upshot is a more active portrayal of perception: perceptual experience is the result of system's current best hypothesis about the cause of sensory stimulation.

Inferential perception is often thought to be Bayesian, (Clark, 2015; Hohwy, 2013): the probability of a given perceptual hypothesis is dependent upon the system's assessment of the probability of the current hypothesis independent of the current sensory evidence[8].

*Bayes' Theorem: P(A|B) = P(B|A) * P(A)/P(B)*

These Bayesian calculations, or something that roughly approximates them, are thought to be occurring unconsciously within the perceptual processing system itself. For example, whether the sensory stimulation that the system has just received was caused by a bee buzzing across my visual field is dependent upon what the system considers the probability that such an event would occur independent of the current stimulation.

When viewing a landscape in the dead of winter where the plant-life is covered in snow and icicles are hanging from trees, the 'bee hypothesis' may not be given much weight compared to if I were standing in a lush field, surrounded by many flowering plants with the sun shining – a scenario in which the perceptual system may rate insect events as having a high probability. Sensory stimulation, of course, is still playing a significant role in the generation of percepts, but so are stored models of the world containing probability density functions: a set of assigned probabilities for different perceptual hypotheses. Stored expectations take the form of *prior probabilities* (priors, for short). Call stored priors within the perceptual system *perceptual priors*.

Perceptual priors and their role in unconscious inference provide a way to make sense of illusory experiences. In cases where one experiences an illusion, one misperceives the world because of failure on the part of their perceptual system to assign proper probabilities to perceptual hypotheses.

[8] I cite Predictive Processing theorists that utilize Bayesian models to make sense of perceptual processing. However, I want to be clear that perceptual processing can be inferential without being Bayesian. Moreover, Bayesian perception need not fit the predictive coding model. I am sticking with the Bayesian model for the purposes of this paper. It need not be true for the basic structure of the model to remain intact.

Recall that people seem to experience images of grey bananas as slightly yellow, which is supported by their tendency to overcompensate whenever asked to make the bananas greyscale by shifting the color too far in the opposite direction. This phenomenon can be explained in terms of perceptual priors. That a banana is yellow has an encoded likelihood within the visual system. If the perceptual prior that the banana is yellow is high enough, then the posterior probability of the perceptual hypothesis that the banana is yellow will beat out the hypothesis that it is grey. The idea is that this is exactly what occurs in the experiments where people fail to change the color of the banana to grey: when it is actually gray, it still looks yellow. This is because the perceptual hypothesis that the banana is yellow is given a higher posterior probability than the hypothesis that it is grey because of the high perceptual prior that a given banana is yellow. To use Siegel's phrase, it's as if perceptual experience is "hijacked" by stored perceptual priors, causing misperception, however slight.

Let's briefly see how this process might play out with idealized equations and value assignments.

o = sense data from the grey banana $P(H_y|o) = P(\text{yellow}) \times P(o|H_y)/P(o)$

H_y = prior that a banana is yellow

H_g = prior that a banana is grey $P(H_g|o) = P(\text{grey}) \times P(o|H_g)/P(o)$

If $P(H_y|o) > P(H_g|o)$, then one will experience the banana as yellow. To illustrate, assign .95 to the perceptual hypothesis that a given banana is yellow, .1 to the occurrence of incoming sense information indicating that the banana is grey, .1 to the sense information would occur given the yellow banana hypothesis, and .9 to the event that sense information would indicate

that the banana is grey given the grey banana hypothesis[9]. The result is that $P(H_y|o) = .95$ and $P(H_g|o) = .9$, which would mean that the perceiver would have the experience of seeing a yellow banana despite the banana being grey.

Bayesian models of perception assume what I have called the Perceptual Resolution Principle – perceptual processing will tend to yield a determinate percept. These theories suggest that even though probabilities are computed and assigned to different perceptual hypotheses, perceptual experience is always the result of the selection of the best current hypothesis, which will rarely, if ever, be something indeterminate.

3.4 The model applied

Let us apply this picture to the Payne's Weapon Identification Task. The Illusory Percept Model states that weapon misidentifications are the result of illusory experiences where the participant has the experience of seeing a gun. Mistakes are influenced by the appearance of a black face before the hand tool display. This is because the participants' 'gun prior' and 'hand tool prior' differ depending on the ethnicity of the face prime. 'Gun prior' and 'hand tool prior' refer to the prior probability that each object would be depicted on the screen in front of participants. For one to have an illusory experience where a hand tool appears to be a gun, the gun prior must be high enough to render the perceptual hypothesis that the object is gun higher than the hypothesis that it's a hand tool.

o = sense data from the socket wrench $P(H_g|o) = P(H_g) \times P(o|H_g)/P(o)$

H_g = prior that an object is a gun

[9] These value assignments are necessarily meant to be realistic. I am using them only to illustrate the point that perceptual hypotheses that are inaccurate may win out over those that are accurate because of the value given to stored priors.

H_{sw}= prior that an object is a socket wrench $P(H_{sw}|o) = P(H_{sw}) \times P(o|H_{sw})/P(o)$

If $P(H_g|o) > P(H_{sw}|o)$, one will see a gun even though there isn't one. The Illusory Percept Model states that the face primes strongly influence the process of perceptual inference – so much so that participants have the illusory experience that a hand tool it is a gun, however briefly. In other words, prior probabilities are assigned to each component of calculations to yield the result that H_g is most probable. Notice how Bayesian inferential perception fills in the details of the previously mentioned Perceptual Resolution Principle. It makes sense of how a determinate representation may be created even though it misrepresents the world. If perceptual processing is constructing representations based on stored information and incoming sensory signal, we should expect it to select the most probable cause of proximal stimulation from a set of possibilities rather than generating something unclear.

The Illusory Percept Model implies that perceptual priors are influenced by ethnicity. There are some findings that provide reason to think that this occurs. In a set of experiments by Eberhardt et al. (2004), an image of a crime-relevant object (either a gun or a knife) was shown to participants, at first being severely degraded so that it was nearly impossible to discern. A series of the same image, but progressively less degraded, was shown to each participant until they were able to discern what was depicted. Before being presented with the images, participants were primed with either a black or a white face. The primes were subliminal, meaning that they were flashed quickly enough such that participants could not consciously recognize that they had seen the faces. Participants that were primed with a black face were able to discern crime-relevant objects (e.g. guns, knives, etc.) at a lower threshold, meaning that they required less images, than those that were primed with a white face. Thus, the black face prime seems to influence perceptual processing in such a way that it causes faster object discrimination.

These experiments provide evidence for the claim that primes influence perceptual processing. If perception operates inferentially, then this experiment provides support for my claim that the black face prime modulates prior probabilities, explaining the effect on the threshold at which participants detected the stimulus. The prime increases the probability for perceptual hypotheses that a crime-relevant object is present in the world, explaining why the threshold for detection of a gun is lowered. Likewise, this model states that in Payne's experiments, primes modulate the 'gun prior'.

In sum, the Illusory Percept Model states that participants in Weapon Identification Tasks err because they have the illusory experience of seeing a gun. Their perceptual systems render the possibility of the presence of a gun more likely whenever they are primed with a black face. If these experiments have any ecological validity, then we should expect that in some real-world cases where police officers think that they've seen a gun in the hand of a black man, it is because of the "hijacking" of experience, they 'see' a gun due to their biased perceptual processing.

4 DECIDING BETWEEN THE TWO MODELS

In this section, I argue that we ought to reject the Executive Failure Model. Then I consider some objections to the Illusory Percept Model and some empirical challenges. I suggest routes for further investigation that may shed light on the mechanisms that cause what I believe to be racially biased perception. Finally, I conclude that the Illusory Percept Model is to be favored as our best current explanatory model for the results of the weapon identification tasks.

4.1 Overriding perception?

The Executive Failure Model states that participants in the Image Experiments correctly perceive the world. Nevertheless, the impulse generated by one's automatic stereotype process causes them to report that the world is other than it is. While the workings of the mind are often

counterintuitive, this implication is striking. The task is to identify an object based on visual information and in Image Experiments, objects are displayed on a screen so that they are clearly discernable. Yet, participants fail at their task because of an interfering mental state, one generated by a process that takes irrelevant information from the world into account, such as the ethnicity of a face presented before the stimulus-to-be-identified is displayed. The pull of involuntary associations is strong enough to mis-report the identity of an object. I question whether there is precedent to suppose that behavior could so easily be hijacked whenever other cognitive processes – perceptual processing in this instance – could better assist the entire system in completing the task at hand. If there is precedent, I'm skeptical that it can generalize to the case of object misidentification. This would be like being told to identify the object in the hand of the next person you see; you clearly see a plumber holding a banana and promptly report "wrench!", or painter holding a cell phone only to report it as a paint brush[10]. The Executive Failure Model comes at this cost, and it seems intuitive that we do not find ourselves pulled by stereotypes so as to misidentify objects as this model suggests.

The authors rely on the analogy of the Stroop task to make sense of the interfering process. At first glance, this analogy seems apt, but it is not. Recall that in the Stroop task, participants see the names of colors printed in ink that are sometimes inconsistent with the color to which the word itself refers to. In the Stroop task, subjects get tripped up in attempting to report another feature of the object. While trying to report the color of the text, another salient and conceptually similar feature of the object, the word that its letters form, may accidentally be reported. In the case of identifying the object in the Image Experiments, participants aren't

[10] These scenarios become significantly more plausible if one thinks that the percept, rather than being clear, is ambiguous, prompting post-perceptual processes fill in the missing details influencing a stereotype-driven report. I consider a model that takes this shape in section 4. The Executive Failure Model, however, takes it that the percept is clear.

reporting another feature of the target; they are reporting the incorrect identity of the object. Whenever a participant reports that a socket wrench is a gun, there is not a gun in their visual field. The content of perception does not contain a second feature that may accidentally be reported as in the Stroop Task – what is perceived is simply a socket wrench. Instead, according to the model, there is an additional associative process where subjects pick out a different feature of the target (or the prime), the man's blackness, and associate it with something that is entirely absent (violence, aggression, weapons, etc.).

It's also critical that the proponents of the Executive Failure Model concede that perception must be playing some role in generating that a hand tool is a gun. If the hand tool were instead a bouquet of flowers or a balloon, I suspect that we wouldn't see misidentifications at the same rate. Thus, the fact that a socket wrench has a similar structure to a gun is significant. If it's playing a causal role in generating the judgment that, according to the Executive Failure Model, is the result of a stereotyping process, what exactly is the role? It's important that the proponent of such a model specify exactly how the percept factors into the misidentification. As it stands, perception's role is unaccounted for.

You might expect stereotype-driven misidentification to occur if perceptual ambiguity arises. My perceptual system is unable to discern the identity of the object, so post-perceptual judgment draws on prior knowledge through stereotypes to label it as a gun. This is plausible, as Stokes and Payne rightly suggest (2010), and as I entertain further in section 5. However, this is hard to square with the fact that objects are presented to subjects very clearly, however brief. Additionally, in follow up experiments, Payne and colleagues allowed participants to alter their responses in case they thought that they were initially wrong (2005). Participants that mistakenly labeled hand tools as guns almost always corrected their answer. If the hand tool were simply

ambiguous and stereotyping led to the judgment that it was a gun, then we are left to wonder how participants were aware that they were incorrect.

My account overcomes these challenges. The Executive Failure Model suggests that misidentification of objects will occur because of our tendency to stereotype associatively, which is at least a counterintuitive and striking implication if we're to believe that this generalizes to everyday cases. An upshot of the Illusory Percept Model is that we will misperceive objects because of the detection of apparent statistical regularities encoded by our cognitive systems. Of course, we often do misperceive the world, so there is nothing strange about generalizing from the Illusory Percept Model to everyday situations. Finally, The Executive Failure Model has no account of how perception factors into generating misidentifications when it's clear that it must factor in somehow. My account, on the other hand, locates the locus of the error in perceptual processing itself.

4.2 A challenge: corrected reports

Payne et al. (2005) attempted to rule out the hypothesis that object misidentification is the result of an illusory perceptual experience by giving participants a chance to correct their answer. After running the basic experiment of instructing subjects to select gun or non-gun, presenting subjects with a face prime, the object to be identified, and asking for a response, the object was immediately masked so that it could not be viewed any longer. Participants were then given an opportunity to change their response without any time pressure. If participants wrongly reported that the object was a gun, prompted with the follow up "actually gun or tool?", they almost always correctly changed their response to tool. That the subjects corrected their initial report at least rules out that the mistakes are cause by a *persisting* illusion. However, as acknowledged by Payne and colleagues (2005), this observation does not rule out that a *fleeting* illusion occurs.

That is, when shown a hand tool, it's possible that subjects briefly – on the order of milliseconds – have the experience of seeing a gun, incorrectly reporting it as such. Further perceptual processing after the mask appears reveals to the subject that they were mistaken, allowing them to report their mistake.

Recall the alternative Executive Failure Model, which rejects the occurrence of a perceptual error and instead says the failure of the executive function to prevent stereotyping to influence behavior is the cause of error. The fleeting illusion hypothesis is at least as plausible as the alternative explanation – the data don't rule decisively in favor of the Executive Failure approach. Here, I present another possible interpretation that fits with the Illusory Percept hypothesis.

Participants are primed and so their visual systems are tuned to anticipate a gun. In the milliseconds during which the participant views the object on the screen, visual processing generates a representation of a gun, causing the subject to report 'gun'. However, as further processing occurs, the brain's hypothesis updates, causing the subject to correct their mistake if prompted. Why, a skeptic might ask, would the brain update the response without further incoming sensory information? If the gun hypothesis was selected, what reason would it have to update?

First, there is the basic fact that neural processing takes time[11]. For their results to count toward the data, participants had to respond to the stimuli before 500ms. The Prediction Error Minimization framework, for instance, suggests that visual processing is an interplay between higher and lower predictions at many different levels throughout the visual pathways in the brain (Clark, 2015). Higher order predictions about the features of the object along with its identity are

[11] Schendan, H. E., & Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *Journal of Cognitive Neuroscience*, *15*(1), 111-135.

being corrected by lower predictions until the perceptual hypothesis that best explains away any discrepancy between prediction and input is generated. Nothing precludes that the process of updating perceptual representations continues to take place for some time after a stimulus is no longer present. If this basic picture is right, the phenomenon of correcting a previously incorrect representation should be expected so long as attention is sustained. Given that subjects are still engaged in the task of identifying the object and are being prompted to sustain attention to correct or confirm their given answer, it is reasonable to suppose that perceptual processing would proceed, however briefly. Thus, the interpretation of the corrected reports is at least underdetermined.

4.3 The verdict

The Illusory Percept Model should be favored over the Executive Failure Model. The latter leaves us wondering why we don't often mis-report the identity of objects due to associative processes whereas the former fits nicely with the banal observation that we often misperceive the world. Moreover, the Executive Failure Model is silent with regards to the role of perceptual representation whereas it's clear that it is performing some function in the behavior observed. My model, moreover, can explain the data generated by Payne and colleagues' corrected report experiments (2005) at least as well as the Executive Failure Model. Finally, the Illusory Percept Model fits nicely with current theories of perception.

There remains another possible interpretation of the data. One might be convinced that the Executive Failure Model's implication that accurate perception is overridden by an independent process is implausible. However, one might suspect that stereotyping judgments occur because the content of perception is unclear. Rather than perceptual processing producing a representation of a hand tool or a gun, perhaps it produces an ambiguous representation,

leaving it up to post-perceptual judgment to decide what's in the world. In the next section, I develop a sketch of the architecture consistent with the claim that perception processing leads to an ambiguity as to whether the object in one's visual field is a hand tool or a gun, causing biased reports as to the identity of the object.

5 THE AMBIGIOUS PERCEPT MODEL

There remains another plausible sketch of the way that the mind generates the observed bias against black men, which locates that error at least partly at the post-perceptual level. It's possible that visual processing generates an ambiguous percept that is neither a gun nor a tool. When the object is misidentified in the Image Experiments, some other mental process distinct from perception judges the percept with an unclear identity to be of a gun. Rather than perception getting the world wrong, it merely finds it to be indeterminate, only for post-perceptual judgment to conclude that the unclear object to be a gun. This model might relate interestingly with the Executive Failure Model: perceptual ambiguity might the exact condition in which we would expect post-perceptual stereotype processing to step in and cause object misidentification (Stokes & Payne, 2010). However, in spelling out the Ambiguous Percept Model, I will leave the exact mechanism that I call post-perceptual judgment to be unspecified, since it might take different forms and for our purposes, not much turns on the exact nature of the process.

There are a few distinct versions of this hypothesis. One possibility is that visual processing systematically outputs an ambiguous percept, perhaps due to how brief the stimuli are presented in the Image Experiments. Another is that bias in the visual system causes what would otherwise be a clear percept of a hand tool to appear ambiguous. Bias in the latter case would be the result differing prior probabilities for the different perceptual hypotheses, as outlined in

Illusory Percept Model. However, rather than one clearly winning out over the other, uncertainty in the system results in ambiguity rather than a clear winner.

Thus, we may have i) ambiguous percepts generated due to a lack of sufficient viewing time of the stimulus and object misidentification being the result of biased post-perceptual judgment, ii) ambiguous percepts generated due to biased priors with no post-perceptual bias, resulting in chance selection of the correct object. In this case, we would still see biased results because of ambiguity arising only or mostly in cases where biased priors generate conflict between perceptual hypotheses. Finally, the last hypothesis states that iii) ambiguous percepts arise due to biased priors and where object misidentification is the result of biased post-perceptual judgment about ambiguous percepts.

What reason do we have to favor the Illusory Percept Model over each version of what I am calling the Ambiguous Percept Model? One might be tempted to argue against the occurrence of ambiguous percepts by appealing to normal phenomenology; our experience of the world isn't often ambiguous and unresolved. Instead, we see definite objects. Even in cases of misperception, we find ourselves surprised when an object is other than it initially seemed – the shadow in my periphery sure seemed as if it were a cat for a moment, but after further inspection, I see that it was only my shadow. It is not often that one has an experience that makes them think "what exactly am I looking at?", and if it does occur, it strikes us as extraordinary. So, one might suggest, we have little reason to suppose that ambiguous percepts regularly occur.

Even if the above claims about normal phenomenology are right, I think this line of reasoning fails. If ambiguous percepts occur, even regularly, we should expect that they would not be encoded into memory. Perceptual ambiguity, if it arises, might do so for objects that are not the focus of our attention. If an object in my periphery is unclear but also unimportant, my

perceptual system need not do any work to bring about a resolution and the experience of such ambiguity would not be striking nor would it be important enough for memory systems to encode it. Perceptual ambiguity might arise for an object that very much is an object of my attention. However, if it arises, further inspection through visual saccades or through other action would cause the ambiguity to resolve quickly. I hypothesize that if this does occur, the experience of ambiguity would not be remembered. Thus, an appeal to phenomenology to refute that idea that perceptual ambiguity is commonplace fails.

Importantly, I do not wish to deny that perceptual ambiguity occurs. If, while looking straight ahead, I hold a previously unseen polaroid photograph up in the periphery of my visual field, I may be able to report some colors and shapes, but I cannot report to you what exactly the photo depicts. In other words, it is ambiguous to me what is causing the impinging sensory information. However, I do wish to claim that perceptual ambiguity will not frequently arise whenever viewing conditions are clear and when the object in question is strongly attended to, as is the case in the Image Experiments. This is an important point because often, viewing conditions might not be clear in real-world scenarios in which police officers must make quick decisions. In such scenarios it's possible that they are relying on post-perceptual judgment. So, to be clear, I do think that perceptual ambiguity might arise in some circumstances.

I believe, however, that there is a strong argument against the regular occurrence of perceptual ambiguity in cases where viewing conditions are clear and when the viewer is attending to the object in question – which one might think accurately describes the conditions of the Image Experiments[12]. First, recall that participants, after misidentifying hand tools as guns, corrected their reports when given the opportunity. If ambiguity called for judgment to identify

[12] Surely, it is assumed that participants are attending carefully to the stimulus, but one might question whether viewing conditions are clear if they only get to view the stimulus for less than a second.

the object's identity, then there would be no reason to expect participants to change their response. This suggests that the object was *not* ambiguous but instead discernable.

Next, recall the Perceptual Resolution Principle. Above, I've argued that perception will tend toward generating unambiguous percepts. I provided experimental support in the form of the binocular rivalry phenomenon. Here, I will extend the degree to which this phenomenon supports the Perceptual Resolution Principle further. Recall the alternating house and face. There is an even more complicated version of binocular rivalry, discovered by Emilio Diaz-Caneja. If images are cut in half so that each eye sees a circle that is composed of two semi-circles, one of which as parallel lines, the other containing concentric circles, rivalry occurs between two coherent circles that either contain only parallel lines or only concentric circles (Howhy, 2013; Diaz-Caneja,1928).

Another experiment by Logothetis and colleagues (1996) shows that normal rivalry persists even if the stimuli presented to each eye are switched every couple of seconds, so that if one is perceiving a house, currently presented to the left eye, even if the image is switched so that the left eye is confronted with a face, the participant will still perceive a house. These variations of binocular rivalry experiments lend support to the idea that discrete, unambiguous perceptual hypotheses are favored over ambiguity. Thus, the Perceptual Resolution Principle, with its independent support, shows us that the Illusory Percept Model is to be preferred over a model that would instead suggest that participants err due to ambiguous perceptual experience. While the basic principles underlying the Ambiguous Percept Model might explain some cases in which perceptual ambiguity arises only to be resolved by post-perceptual judgment, I believe that we should reject that it is the best explanation for the Weapon Identification Tasks.

6 FUTURE EMPIRICAL CONSIDERATIONS

I've argued that my proposed Illusory Percept Model ought to be favored on the grounds that it accords well with leading theories of perception and because its competitor leads to implausible consequences. However, there aren't any decisive empirical reasons to favor one over the other. Are there any ways to test the differing hypotheses?

It's possible to identify the areas of the brain responsible for representing particular stimuli. In fact, binocular rivalry (see section 3.1) has been used to identify the neural correlates of conscious experience (Wu, 2018). Using fMRI, we might analyze brain activation while participants view a binocular rivalry stimulus with one half being a gun and the other a hand tool such as a socket wrench. We can ask the subject to report whenever they are seeing a gun and whenever they see the socket wrench. Correlating their reports and examining the difference between activation should allow for the localization of neural realizers for each experience. We can then have subjects complete the Image Experiment task in the fMRI. Whenever they mistakenly report that a hand tool is a gun, we can examine brain activation. If we succeed in localizing the area in the cortex that correlates with conscious perception of a gun as determined in the binocular rivalry imaging experiment, we can examine whether the same activation occurs whenever participants mistakenly report that the socket wrench is a gun. If so, then we have strong reason to think that their experience was of a gun as opposed to the possibility that they were correctly seeing the socket wrench but reporting that it was a gun due to other mental states.

7 CONCLUSION

Until further empirical investigation is conducted and provides evidence for or against the interpretations of the current data, the Illusory Percept Model should be favored over the Executive Failure Hypothesis and the Ambiguous Percept Model. Current leading theories of

perception suggest that processing occurs inferentially, meaning that the brain is inferring the cause of current sensory stimulation. This leads to the possibility that the brain selects a perceptual hypothesis that is incorrect based on stored prior probabilities. These theories make sense of illusory phenomena, providing a mechanism that explains how participants in weapon identification tasks may be misperceiving the world.

If the Illusory Percept Model really does explain some instances of racial bias manifesting in behavior, questions arise regarding effective training methods. It's important to note that while the term 'implicit bias' is often associated with non-perceptual doxastic states, the concept as a general category undoubtedly subsumes perceptual bias. We can look to the vast literature on implicit bias training to examine potential effect training methods. In fact, Payne (2001) reveals that prompting participants to control for bias moderates the observed effect of bias in the Image Experiments. This is consistent with the Eberhardt priming experiments mentioned above, in which participants that were subliminally primed with a black face were quicker to discern noisy images of crime-related objects (2004). The activation of certain concepts may have an effect on the relative weights given to different perceptual hypotheses. If this is the case, then lessening activation by weakening associations may be a way to reduce the likelihood of misperception. Lessening the activation might be achieved through many of the investigated ways to reduce implicit bias, such as counterstereotypes, where people are exposed to stereotype *dis*confirming individuals (Fridell, 2017). Counterstereotype exposure might lessen the association between concepts, their concurrent activation, and the effect on perceptual priors. Further theoretical support for the idea of counterstereotypes being a solution comes from the observation that exposure to stimuli that violated encoded visual expectations shaped perceptual priors. This is demonstrated by Adams et al. (2004) where the researchers show that the light-

from-above prior can be changed, at least temporarily, demonstrating the malleability of probabilities deployed in perceptual inference.

Philosophers need to assess what racially biased perception means for epistemology and for notions of moral responsibility. If the world can be systematically misrepresented due to expectation-driven perceptual processing, then perception may not be as reliable as we often take it to be. The idea that populations might have their visual systems tuned such that they generate racially biased misrepresentations poses an epistemological problem. Are we to hold people responsible for having racially-biased experience? And more broadly, when are we not justified in trusting the content of our experience?

The first question turns on the degree to which individuals have control over the process of perceptual inference. Holding someone responsible for their experience is surely a strange suggestion. However, when such experiences arise systematically, are potentially avoidable, and may lead to someone's death, the question must be entertained. Fortunately, deleterious actions based on brief illusions aren't likely to be ubiquitous, but in cases where such actions can occur and have dire consequences as in the case of police officers firing their weapons, we must consider whether individuals can be held responsible. If further empirical investigation reveals that illusory experiences are generated by bias and that training can re-tune perceptual inference to reduce the likelihood of such experiences, then it seems to me that one's disposition to have certain experiences is morally appraisable. In a country where police killings are pervasive, it is incumbent upon individuals and institutions that have power to take the lives of others to be sure that racial bias plays no role in their dealings with citizens, especially whenever such killings disproportionately affect the groups that are victims of such bias. The existence of effective training methods that are likely to reduce bias generates an obligation for the institutions to be

sure that police officers take such steps to reduce prejudice in their actions, beliefs, assumptions, and perceptions.

Susanna Siegel has defended the thesis that experiences are epistemically appraisable (2016), arguing as that inferences generally are subject to epistemic assessment and if perception proceeds by inference, then we can hold perceptual experience appraisable. When isn't perceptual experience justified? How can we know whenever an experience is the result of poor inference? Investigating the role that imagination and mental imagery play in perception, Bence Nanay calls for a naturalized epistemology that examines the reliability of processes that give rise to mental imagery (2018). Illusions of guns due to poor perceptual inference is an instance of mental imagery corrupting the epistemic role of perceptual experience. Under what circumstances does perceptual processing get things wrong because of encoded perceptual priors? To answer such questions, the mechanism by which perceptual priors are tuned must be uncovered. I suggested earlier that tuning might occur within the visual system, insulated from one's beliefs. But priors might also be influenced due to the cognitive penetration of perceptual processing. Elucidating these mechanisms will aid in understanding how to assess the reliability of perceptual experience – an ostensibly reliable process that is prone to the occasional, but sometimes detrimental error.

To summarize, I've constructed and put forth a theory where racially biased object misidentification is the result of illusory experience. Importantly, I believe to have only identified one way in which bias might arise. Further empirical work can reveal to what extent racially biased perception is a problem and further philosophical work can aid in elucidating its consequences for epistemology and responsibility.

REFERENCES

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the ‘light-from-above' prior. Nature neuroscience, 7(10), 1057.

Alais, D., O'Shea, R. P., Mesana-Alais, C., & Wilson, I. G. (2000). On binocular alternation. *Perception*, *29*(12), 1437-1445.

Bremner, A. J., Doherty, M. J., Caparos, S., de Fockert, J., Linnell, K. J., & Davidoff, J. (2016). Effects of Culture and the Urban Environment on the Development of the E bbinghaus Illusion. Child development, 87(3), 962-981.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. Psychological bulletin, 86(2), 307.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of personality and social psychology*, *83*(6), 1314.

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: a sociofunctional threat-based approach to" prejudice". *Journal of personality and social psychology*, *88*(5), 770.

Diaz-Caneja, E. (1928). On binocular alternation. *Ann. Ocul*, 721-731.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing Black: Race, crime, and visual processing. Journal of personality and social psychology, 87(6), 876.

Fridell L. (2017) The Science of Implicit Bias and Implications for Policing. In: Producing Bias-Free Policing. SpringerBriefs in Criminology. Springer, Cham

Greenwald, A. G., Oakes, M. A., & Hoffman, H. G. (2003). Targets of discrimination: Effects of race on responses to weapons holders. *Journal of Experimental Social Psychology*, *39*(4), 399-405.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature neuroscience*, *9*(11), 1367.

Hatfield, G. (2002). Perception as unconscious inference. *Perception and the physical world: Psychological and philosophical issues in perception*, 113-143.

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. Progress in neurobiology, 84(4), 405-439.

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. Journal of personality and social psychology, 57(2), 165.

Macpherson, F. (2017). The relationship between cognitive penetration and predictive coding. *Consciousness and cognition*, *47*, 6-16.

Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. Noûs, 50(3), 629-658.

Nanay, B. (2018). Multimodal mental imagery. *Cortex*, *105*, 125-134.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of personality and social psychology*, *81*(2), 181.

Payne, B. K. (2006). Weapon bias: Split-second decisions and unintended stereotyping. Current Directions in Psychological Science, 15(6), 287-291.

Payne, B. K., Shimizu, Y., & Jacoby, L. L. (2005). Mental control and visual illusions: Toward explaining race-biased weapon misidentifications. *Journal of Experimental Social Psychology*, *41*(1), 36-47.

Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science*, *357*(6351), 596-600.

Sagar, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in Black and White children's perceptions of ambiguously aggressive acts. *Journal of personality and social psychology*, *39*(4), 590.

Siegel, S. (2016). The rationality of perception. Oxford University Press.

Stokes, M. B., & Payne, B. K. (2010). Mental control and visual illusions: Errors of action and construal in race-based weapon misidentification. *2010*, 295-305.

Sun, J., & Perona, P. (1998). Where is the sun? Nature neuroscience, 1(3), 183.

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of personality and social psychology*, *113*(1), 59.

Wu, Wayne, "The Neuroscience of Consciousness", The Stanford Encyclopedia of Philosophy (Winter 2018 Edition), Edward N. Zalta (ed.), forthcoming