Bias in a Biased System: Visual Perceptual Prejudice

Jessie Munton

October 2017

Abstract: The visual system responds to both an excess of available information and an underdetermination problem by relying on priors that facilitate its successful navigation of an environment on the basis of past encounters. These priors can concern demographic features, such as race or gender, in a manner that resembles prejudicial bias. This focuses our attention on the following problem: how can we identify prejudicial bias within a system that relies on a kind of structural bias to accomplish its goals? Taking two bodies of recent empirical work as case studies, I explore and reject three possible criteria for demarcating problematic instances of bias. I identify a principled reason why standard epistemic criteria cannot accomplish the task. I instead propose an adoption of a skill-based model of visual perception that allows for multi-dimensional evaluation relative to a set of potentially competing goals.

0. Bias in a biased system

The term 'bias' is used in a variety of ways. Two uses are of particular relevance to this chapter. We often use 'bias' colloquially to mean a kind of prejudice, against a particular social group. This is the sense in which we talk of "implicit bias" grounding discrimination, or of a system that is unfairly "biased" against a racial minority, for instance. This sense of bias is inherently freighted with a negative valence: it has problematic ethical upshots. In addition, it is often taken to be epistemically flawed, either because it is grounded in accurate information or irrational cognitive processes, or because it perpetuates flawed reasoning downstream of it.[1] We can call this kind of bias *prejudicial bias*. Prejudicial bias generally concerns demographic groups: people of a particular race, gender, sexual orientation, or class.

But in science more generally, bias also has a thinner, formal sense, meaning any kind of weighting in a testing process that systematically skews the outcome. This is what we have in mind when we talk of a confirmation bias, for instance, when a process is skewed towards producing results which confirm a previously endorsed hypothesis. Bias in this second sense may sometimes also be prejudicial against particular demographic groups, but it need not be. And some kinds of action recommended to combat the first, negatively valenced kind of bias may be instances of the second, formal sort of bias, such as deliberately assigning readings by female authors on a particular topic to combat their underrepresentation on philosophy syllabi, for

[1] For a range of philosophical perspectives on the nature of implicit prejudicial bias and the sorts of problems it can give rise to see Brownstein and Saul (2016). For a discussion of the downstream epistemic problems that prejudicial bias can give rise to see Gendler (2011).

instance, or practices of affirmative action. We can call this second kind of bias *formal bias*.[2]

Formal bias need not always constitute an epistemic flaw, though it may sometimes do so. And under certain conditions, it can constitute an ecologically rational response, one that maximizes the quantity and quality of information at the disposal of an organism. It is not, therefore, inherently epistemically valenced. Nor is it inherently morally valenced. In this sense, it is a neutral category.

What is the relationship between prejudicial bias and formal bias? Is prejudicial bias just an instance of formal bias that happens to have a particular demographic orientation, or is it a category that is marked out itself by further structural flaws in reasoning, for instance? This question has important upshots for our understanding of the epistemic status of prejudicial bias: are the epistemic problems that frequently seem to accompany it merely *contingent*, or is it in fact distinguished by structural properties which are themselves constitutive of an epistemic flaw?

Standard accounts of epistemic normativity have tended to abstract away from the particular content of the beliefs or arguments under evaluation, in favor of focusing on formal requirements they must meet to attain a certain standard, such as knowledge, or justification. Evidentialism for instance, claims that believers must conform their beliefs to the evidence. This requirement applies irrespective of the *subject matter* of the beliefs in question (Connee and Feldman 2004). Similarly, reliabilism about knowledge or justification emphasizes the significance of accuracy or the truth-ratio of a given means of producing belief (Goldman 1979; Kornblith 2002). This is again a formal requirement: it does not pertain to the contents of the beliefs under scrutiny. We shall call a system of evaluation that can be described and applied without appeal to content a "formal" method of evaluation.[3] If prejudicial bias is epistemically flawed *as a category*, then we would expect it to be distinguished not just by its demographic focus, but in addition by certain structural features in virtue of which those flaws arise, since standard methods of epistemic evaluation are not sensitive to whether or not the content of belief concerns a demographic group.

In this chapter I pursue this question, of what distinguishes prejudicial bias from the broader class of formal bias, specifically in the context of the visual system. Visual perception is a process which demonstrates systematic formal bias. It relies on previously-encountered information to guide its accumulation and interpretation of new data. There is also evidence that visual perception can demonstrate a sensitivity to demographic categories including race and gender, in a manner that resembles a

[2] See Greenwald and Krieger (2006) for a related distinction between neutral "response bias" and implicit, discriminatory bias.

[3] The exception to this general rule is the literature on pragmatic encroachment, the view that what a subject has practically at stake can "encroach" on epistemic standards applied to them (Fantl and McGrath 2002; Hawthorne 2003; Stanley 2005; Weatherson 2012). Higher stakes make it harder to *know* a given proposition, for instance. Applying this variable standard requires attention to the contents of the belief and details of the believer's situation.

kind of prejudicial bias.[4] This throws up the following question: how are we to identify instances of prejudicial bias within a formally biased system? I argue that there are cases of prejudicial bias which are indistinguishable in formal terms from the standard operation of the visual system. But I also claim that these cases do manifest distinctive epistemic flaws.

Understanding how we can square this circle requires a reappraisal of our epistemic evaluation of visual perception more generally. To capture the full range of instances of visual prejudicial bias we need to conceptualize visual perception as an active practice that develops over time, that crucially involves not just transitions between pieces of information but decisions about *what information to acquire,* and that can consequently only be fully epistemically appraised relative to a goal and an environment. The introduction of these additional parameter gives us the room we need to individuate prejudicial bias in a way that reveals it to be constitutively tied to a distinctive kind of epistemic flaw. In doing so, we make some progress on the broader question of how to identify prejudicial bias within a structurally formally biased system.

This chapter proceeds as follows. In the first section I describe how the visual system is both formally biased at a structural level, and capable of engaging in a kind of prejudicial bias. I draw on two bodies of empirical work - on face perception, and the resolution of perceptual ambiguity - to make this point. In the third section I consider candidate criteria that fail to demarcate prejudicial bias from mere formal bias. In the final section I offer an analysis of *why* these criteria are bound to fail, and propose a different approach to visual perception, one that emphasizes its status as a skill honed through active practice in a given environment, whose evaluation is only possible relative to a set of goals.

1. Vision as a biased system

The visual system faces two major challenges. On the one hand, it has too much information: the information available from the environment at any one moment far surpasses what the brain can process (Summerfield and Egner 2009). On the other hand it has too little information: the 3D world is projected onto the 2D retina in a manner that is essentially ambiguous, and the brain must work to reconstruct a representation of the environment from that impoverished data (Scholl 2005). The solution to both of these problems is similar: the visual system relies on information it has encountered in the past, encoding previous environmental regularities, to let it select those pieces of information most likely to issue an appropriate epistemic reward, and to help it resolve indeterminacy in retinal data. In some sense, the visual

[4] We tend to think of prejudicial bias as something that manifests in *cognition,* and that is, if not the upshot person-level processing, at least something for which an individual may be blamed or held accountable in some sense. I am using prejudicial bias in a thinner sense here, that can apply to the upshot of any level of processing, realized by an individual, a sub-system of an individual, or even a computer algorithm.

system is thereby systematically biased, biased towards the interpretation of the novel in line with the familiar, thereby winning greater reliability, (as measure by the increased rate of accuracy of the resulting experiences), at the cost of a loss of sensitivity to surprising data.

Bayesian models of visual perception bring out this biased structure. These models propose that the visual system overcomes the uncertainty inherent in retinal stimulation by relying on a set of priors that guide its interpretation of new data, and that it performs these calculations roughly in line with Bayes theorem (J. Feldman 2014; Rescorla 2015).[5] To take a toy example, suppose you are looking at a line. The retinal space occupied by that line is consistent with it being a number of different lengths, depending on how far it is from you, and the angle at which it is positioned. Put simply, the visual system can resolve that uncertainty by drawing on prior probability distributions over possible values for its length, angle and distance from you. It can then calculate the most likely value for each of those parameters, and in so doing arrive at a more or less determinate representation of that aspect of its environment.[6]

In relying on priors in this way, the visual system does not approach every problem it faces with an open mind, so to speak. If it did, the chances of it accurately finding the relevant information, or resolving the ambiguity, would be lower. By relying on prior information of some kind, it increases its chances of accurately perceiving its environment, and of locating relevant information within it. It also introduces a kind of minimal confirmation bias, since ambiguities in novel stimuli are resolved in line with previous regularities. We do not come at the world with fresh eyes every time. Instead, our perception of the new is colored by our perception of the old.

But note that although Bayesian updating is an instance of a process that is neutrally biased, it can still be optimally rational. In fact, it can provide a counterweight to prejudicial bias, in so far as proper attention to base rates often forestalls the formation of prejudicially biased belief about particular demographic groups. Attending to just how unlikely it is that one would be robbed at all, for instance, makes it irrational to have anything but the lowest confidence that an approaching male is likely to rob you, despite the higher likelihood that a man rather than a woman would rob you, were you to be robbed at all.[7]

---

[5] Bayes theorem provides a formula for integrating new information with old (the "priors"), to arrive at the probability of a candidate hypothesis, conditional on the relevant evidence. It states that the probability of a given hypothesis given a piece of evidence is equal to the probability of the evidence given the hypothesis, multiplied by the probability of the hypothesis ($(p(h \mid e) = p(e \mid h)\, p(h))$

[6] For related examples described in considerably more detail see Kersten and Yuille (2003)

[7] This is where Lee Jussim errs when he writes as follows: "Let's assume for a moment that 30% of motorcycle gang members are arrested for violent behavior at some point in their lives, and 0.3% of ballerinas are arrested for violent behavior at some point in their lives. People who know this are being completely reasonable and rational if, on dark streets or at lonely train stations, they avoid the bikers more than ballerinas, in the absence of much other individuating information about them" (2009:215). Supposing that the rationality of avoiding

The visual system demonstrates this kind of neutral bias not just in the resolution of uncertainty but in the processes of selecting visual input. Past information is similarly drawn on when strategizing *where* to look within a scene to find a given piece of information. Context primes subjects to identify scene-consistent objects: bread in a kitchen, for instance, or a football player on a football field, because those are the contexts in which we have encountered those objects previously (Palmer 1975; Davenport and Potter 2004). It can similarly help us predict where to look within those scenes to find relevant stimuli (Wolfe and Horowitz 2017).

This process of learning to prioritize certain pieces of information gives rise to the development of a kind of perceptual expertise, the development of fine-grained capacities of discrimination and recognition with frequently encountered stimuli (Gauthier, Tarr, and Bub 2009). We can recognize instances of this specialization when reflecting on our own visual experiences: the novice gardener has to inspect each plant sprout carefully to distinguish weed from seedling where the expert gardener has no such difficulty; the radiologist now easily parses x-rays into the body parts they represent when at first they seemed a confusing mass of black and white; the experienced mason immediately spots the imperfect grouting whilst his new assistant cannot see the difference. These effects rely in part on the direction of attention, but also on changes to lower level perceptual processing (Harel, Kravitz, and Baker 2013):[8] the visual system is capable of *changing* in response to previous tasks and encounters. This expertise is once again an instance of formal bias: what information is extracted from a particular scene depends on the system's prior encounters.

This kind of neutral bias in the visual system becomes more worrying when we consider the possibility that the visual system draws on stored information not just about low-level features of its environment such as the likely length and incline of lines, but in addition about higher level features, such as race or gender. Below I briefly outline two different bodies of empirical work in visual perception which suggest this is a possibility: work on face perception, and work investigating how stereotypes influence the speeded identification of objects.

i. Prejudicial bias through facial expertise

Our capacity to recognize and read information from faces is a revealing window on the visual system's sensitivity to race. We know, for instance, that visual adaptation to faces is race specific, suggesting that faces of different races are coded for by different

bikers depends on a credence that they might harm you of greater than .01, given almost any kind of plausible base rate, it remains irrational to avoid bikers just on the basis of their identity, (i.e. the absence of more specific information that they intend to pose a threat to you).

[8] See Siegel (2010) for the claim that cases of expertise similar to these can result in a change in phenomenal contents of the experience.

neural populations.[9] Adaptation to Caucasian faces with artificially distorted features, for instance, does not transfer to Chinese faces, and vice versa. This happens in a manner that suggests a sensitivity not just to low-level physical features but to the social category of the face (Jaquet, Rhodes, and Hayward 2007, 2008). Similarly, adaptation effects are specific to male or female faces (Jaquet and Rhodes 2008; Little, DeBruine, and Jones 2005).

In fact, there is good evidence that our skills at face perception are systematically arranged along racial lines. We are more accurate at recognizing own-race and dominant-race faces, with both fewer false positives and negatives (Meissner and Brigham 2001). A white person in a majority white society will generally be better at recognizing other white faces than black or Asian faces, for instance. A black person living in a majority black society is likely to be less good at recognizing white faces than black faces. Disturbingly, the difference in performance is not limited to recognition: subjects' identification of emotion is also more accurate for same-race faces (Elfenbeim and Ambady 2002).

This differentiation emerges in infancy: Kelly et al. (2007) describe a process of "perceptual narrowing" that emerges over the first 9 months of life and involves a loss of capacity to recognize other-race faces. Whilst 3-month-old Caucasian infants could recognize white, Asian and black faces after a brief period of exposure to color images of them, 6 month olds could recognize only white and Asian faces, and 9 month olds could recognize only Caucasian faces.

Though this bias towards reading faces from one's own or dominant racial groups is well-established, the mechanism behind it is contested. One interpretation is that exposure to certain race faces in effect allows the individual to develop more fine-grained capacities for distinguishing between those faces.[10] Valentine (1991; Valentine, Lewis, and Hills 2016) for instance, offers an interpretation of this "same-race face effect" in terms of a multi-dimensional "face space", within which perceivers encode faces as vectors, depending on their deviation from a prototypical average.[11] Expertise with certain types of faces enlarges certain areas of the face-space allowing for finer-grained distinctions. Exposure to faces of one's own race stretches that area of face space. As a result, similar faces are in effect situated at a greater distance from one another in that space, making them easier to distinguish and consequently more

[9] Visual adaptation effects occur when exposure to stimuli which manifest a particular property results in reduced sensitivity to that property, with the result that the subject may experience illusory after-effects (Clifford et al. 2007). A familiar example is the waterfall illusion: after looking at a waterfall for a period of time, upon looking away one may be subject to the illusion that one's surroundings are moving *upwards*.

[10] An alternative possibility is that the same-race face effect stems from the subject devoting mental resources to "coding" the race of other-race faces (Levin 1996, 2000).For empirical critique of this "race coding" hypothesis see Rhodes et al. (2009).

[11] See too Catz et al. (2009)for a contrasting model of face space along "single dimensions", that avoid appeal to the complex dimensions of race, gender and age. See too Catz et al. 2009 for a model of face space along "single dimensions", excluding complex dimensions of race, gender and age.)

memorable. Rhodes and Leopold (2011) describe this as a process of "norm-based coding": faces are encoded relative to a "norm", the average of faces encountered so far. Their relationship to that norm has significant upshots for what information you can read from them.

The visual system is in effect forced to accept a pay-off, between a capacity to read information from a broad range of faces, and a capacity for a fine-grained read off of information from the sorts of faces it most commonly encounters. In fact, there are multiply pay-offs in play here. We are actually less efficient at categorizing own-race faces by race than we are at categorizing other-race faces. Quicker and more accurate recognition seems to comes at the cost of less efficient categorization. Ge et al. suggest this "antagonistic interaction" reflects a trade-off "between processing individual identity and categorical facial information of faces from own and other races" (2009:1200).

What are the epistemic upshots of this norm-based coding of facial information?[12] There are two relevant noteworthy features of it. One is that it allows for the extraction of more, and more fine-grained, information from faces. You become expert at the sorts of face you most commonly see. That expertise seems like a positive result, epistemically speaking, in so far as you get more information than if you did not develop it. But another upshot of it is that there is a *discrepancy* in your ability to retrieve information about people depending on their race. You recognize some people less readily than others, and read information from their races less easily. This discrepancy itself is significant: it leaves you better equipped to interpret and interact with individuals of one race than another. This information about white people, say, comes at a cost, the cost of racial neutrality, and neutrality itself may be epistemically valuable in certain situations.

One problem this gives rise to is a kind of snowballing effect of the expertise: the dominant race individual extracts more information from same-race faces, thus further expanding those regions of her facial expertise. Moreover, as a result she is better positioned to further interact with those individuals, and to cultivate relationships with them through improved skills at identifying emotions for instance. That in turn is likely to further skew her facial perception skills towards reading same-race faces.

Further difficulties are added when we consider that this failure of neutrality does not advertise itself to the subject of the experience. This is especially concerning when we consider the discrepancy in reading emotion from same and different-race faces. It is unlikely that the individual is capable of detecting that their perceptual skills are uneven in this regard, unless their attention is drawn to it. After the same exposure to faces of two different races an individual will come away with different quantities of information, facilitating better recall and identification of emotions in one case then

[12] For further, rich discussion of the epistemic costs of same-race face effects see Gendler (2011).

the other, whilst being liable to think of themselves as having equal access to information from both faces. That in turn encourages inaccurate inferences about the resulting discrepancy in the information they have: that other-race faces *display* less emotion, for instance.

But this structure of expertise also has a degree of ecological validity: an infant who failed to develop fine-grained skills at reading the faces that surrounded him would undoubtedly suffer in complex social interactions. A similar structure of expertise is, moreover, to be found in a range of other non-racial contexts, as individuals become skilled at detecting stimuli that serve the use to which they put their visual perceptual faculties: reading information from x-rays, or leaves, or livestock. And yet the same-race face effect seems like an instance in which the visual system demonstrates and thereby perpetuates a kind of racial bias. This raises the question: what distinguishes face-recognition as an instance of *prejudicial* bias, against a backdrop of formally biased task-specific expertise within the visual system more broadly?

ii. Prejudicial bias through the speeded recognition of ambiguous stimuli

I turn now to a second way in which visual perception could manifest, and thereby perpetuate, a kind of prejudicial bias against certain demographic categories. This effect springs from the way in which the visual system resolves ambiguity in in-coming data on the basis of priors that encode information about environmental regularities. Could associations with racial categories, or previous regularities, prime you to identify properties stereotypically associated with race? If that were the case, the visual system could in effect recapitulate prejudicial racial biases, interpreting new data in line with previously encountered regularities between different races and certain contexts. Suppose, for instance, that you worked at Yale Law School, where only 8% of the ladder faculty are black or Hispanic. When you come to eat in the dining hall, however, those proportions are roughly inverted amongst the catering and maintenance staff (Yale Office of Institutional Research 2016; National Center for Education Statistics 2016). If your visual system could encode those context-specific regularities, it could interpret novel data in line with them: the ambiguous face at the periphery of your visual field could be resolved in one direction when you are in a faculty seminar, another direction when you are in the dining hall. In allowing for that stereotype-consistent resolution, your visual system would in effect be racially biased, eliding information that contradicted its preexisting racial associations.

The possible impact of these kinds of effect become more disturbing if we countenance the possibility that your visual system is sensitive not just to regularities in the real world, but to regularities encountered in the media, via photographs or film, through television and the internet. This opens the door to a far wider scope of stereotypical associations: consider the regular association of violent crime and young black men through news reporting.[13] The visual system is presumably insensitive to

[13] Take, for example, the findings of a report of the REACH media monitoring project on media representations of black young men and boys in the UK. The report concluded that "In our survey of the mainstream news media, we have clearly identified that black young men and

the accuracy or proportionality of that reporting. It may nonetheless absorb the regular correlation between the two, storing it as a possible prior to draw on in the future selection of relevant information or resolution of noise and ambiguity in its incoming signal. But even if we exclude influences that are disproportionate or inaccurate, we are left with the worry that simply passage through a racist world may be enough on its own to leave the visual system with racially-skewed priors.

It is a disturbing thought that the priming of certain concepts, or a certain pattern of past exposure, could determine what you end up seeing as you look at the world around you, in a manner consistent with racist stereotypes. Could such influences make you more likely to identify ambiguous objects as crime-related, for instance? A variety of recent work in vision science purports to follow through on demonstrating that racial stereotypes may influence visual perception. I will focus here on work done by Joshua Correll and his colleagues, which aims in part to synthesize and adjudicate between previous studies in this area.

Take a representative study by Correll (2007), in which subjects played a crude kind of video game. A series of photos each showing a man in one of various contexts (in front of a car park, on the street, in a park) appeared on the screen before them. In some of the images the man was holding a gun, and in others a harmless object such as a cellphones. The subjects' task was to indicate "shoot" in response to those individuals who are holding weapons and "not shoot" in response to those holding innocuous objects. Subjects responded under time pressure by pressing a computer key to indicate their choice. Correll was interested to see whether the speed, and pattern of errors in their responses were sensitive to the race of the target individual featured in the photo. Their results displayed a kind of racial bias: in line with stereotypes associating black men and crime, subjects were quicker to select the "shoot" response for images of black men holding guns, and were more likely to mistakenly shoot black men than white men. One important question this gives rise to is whether the subjects just *responded* in a manner consistent with racial bias, or whether their visual experience of the stimulus itself was partly responsible for this pattern of results.

Earlier work by Jennifer Eberhardt (2004) suggested that effects of this kind could be genuinely perceptual: Eberhardt tasked subjects with identifying photographs of objects as quickly as possible. The objects were degraded with noise, which cleared frame by frame. Eberhardt found that subjects were quicker to identify crime-relevant objects when primed with a black face than with a white face. In this case it seemed

---

boys are regularly associated with negative news values. …. Young men and boys in general were regularly reported in relation to negative news values, just over 4 in 10 stories being crime-related, while stories about wider social issues, such as education and health, were not as frequently or prominently reported. By contrast, however, close to 7 in 10 stories of black young men and boys related in some form to crime - a comparatively higher figure than in coverage of young men and boys more generally. Indeed, the most significant characteristic of the news media coverage surrounding black young men and boys was their association within news narratives with crime" (Cushion, Moore, and Jewell 2011: 87).

that the different in performance had to be attributable to them seeing the relevant objects sooner, through the noise laid over them.

Work by Keith Payne by contrast suggested that the kind of effect demonstrated by Correll could best be explained not in terms of the impact of racial stereotypes on visual perception directly, but merely on the individual's capacity to control their responses. Keith Payne had initially found that priming individuals with a black face as opposed to a white face made them more likely to identify an image as a gun than as a tool (Payne 2001). But in later work he found that when the time pressure to respond was alleviated, subjects' stereotype-consistent errors evaporated. And subjects could almost always identify when they had made such an error, suggesting that it was a problem at the level of response rather than perception (Payne, Shimizu, and Jacoby 2005).

Correll and his colleagues (2015), interested to resolve these competing interpretations of the influence of racial associations on perceptual experience, performed a "diffusion analysis" on their earlier data. This process uses subjects' accuracy and latency to model their decision-making process. In particular it aims to pull apart the relative significance of three possible points at which the biased pattern of response could emerge. Firstly, it could be that subjects *start* from a position that is already biased towards a particular response tendency, i.e. shooting, when confronted with black targets. Secondly, it could be that they *move* more quickly from their starting point to the point at which they have apparently accumulated sufficient evidence to respond on stereotype-congruent trials. Or, thirdly, it could be that the difference in response is accounted for by a speedier *action* response, after that process of accumulating information has been completed.

Correll found that the difference was in the rate at which subjects "accumulated evidence", that is, the time it took for them to reach a point at which they were willing to make a decision. For armed targets, participants accumulated evidence more quickly when the target was black than when the target was white. This difference in the so-called "drift rate" suggests a perceptual element to the effect, since visual perception was the key means by which subjects accumulated the evidence in question. Correll's verdict was as follows:

"Overall, then, the results from diffusion model analysis suggest that participants accumulate evidence more quickly when targets "fit" prevalent stereotypes, but more slowly or gradually when targets violate those stereotypes. This pattern suggests that the targets' race may guide visual interpretation of the object, perhaps by offering supplemental information" (2015:225).

The same paper also approached the question of how stereotypes contributed to the pattern of biased responses by tracking subjects' eye gaze. Since acuity falls off sharply within a few degrees of the fovea, the most light-sensitive part of the retina, where subjects look tells us what information they are prioritizing whilst undertaking the

task. Since the task was to differentiate the object the target was holding, one would naturally expect that subjects would in all cases look directly at that target object. Surprisingly, Correll found a difference here that depended on the race of the target. The visual angle (the angle between the subject's visual focal point and the relevant object) was significantly greater for blacks than whites, suggesting "that participants shot Blacks with relatively low visual resolution or clarity concerning the object, whereas they achieved much greater visual resolution before shooting an armed white" (2015:227). When the target was black, subjects were attending to other parts of the image, such as the face, rather than looking directly at the relevant object. Correll et al. offered the following rationale for those results: "[i]f race augments the available visual information on [stereotype congruent] trials, participants should require less of the available *objective* information. …If a gun in the hands of a White man somehow looks less readily like a gun, participants … should seek greater clarity through an extended visual search. In essence, they may require more concrete, clearer objective information due to the fact that race impairs subjective interpretation." (2015:225).

The authors' overall verdict is as follows: "This is exactly the pattern we would predict if racial stereotypes augment visual processing, leading participants to more quickly interpret ambiguous evidence, such that they reach a stereotypic decision more quickly (as measured by the drift rate index) and so require less fine-grained information (as measured by the visual angle index)." (2015:228)

Accepting the authors' interpretation of their work, this looks like a case of prejudicial racial bias: the visual system is precluded from gathering relevant information that would counteract a prejudicial association it draws on, by the influence of that very association. In what follows, I will call cases of this sort, in which prejudicial bias arises from a reliance on stored information, "Correll cases". Our condemnation of the structure of influence in these cases is complicated by an appreciation of the systematic reliance of the visual system on stored information, in the manner described above. We know that more generally, objects appearing in a consistent background context (a loaf of bread on a kitchen counter) are identified more quickly than when they appear in incongruous contexts (a drum in the same setting) (Palmer 1975; Davenport and Potter 2004). In these cases too, the visual system draws on a "stereotype", an association between contexts and objects, just as it does in the Correll cases, in which it relies on an association between race and crime. In this sense, the visual system is systematically "biased". What is it, then, that differentiates the prejudicial case from cases of merely neutral bias? Intuitively prejudicial cases are particularly epistemically flawed. What we are ideally looking for, then, is a criterion that distinguish cases of prejudicial bias via an epistemically valenced criterion, thereby explaining our intuitive sense that these cases constitute a distinct, and a distinctively epistemically problematic, kind.

2. Candidate criteria

a. Irrational transformations

It can seem obvious that prejudicial bias, even in the perceptual case, involves *bad reasoning*: the adoption of irrational priors that do not match the available evidence, or a failure to conditionalize appropriately on a prior. Susanna Siegel (2013b; 2016) argues that we can rationally appraise the sub-personal transitions involved in the formation of perceptual experiences. Irrational transitions have a negative impact on the capacity of the resulting visual experience to justify belief, just as irrational transitions in the formation of belief diminish its rational power.[14] We could apply this approach to the evaluation of the subpersonal processes of Bayesian updating posited to obtain in visual perception. Instances of prejudicial bias could be marked out in term of deviant processes of updating. According to Siegel's proposal, this would in turn reduce the epistemic power of the resulting experiences, that is, their capacity to justify belief.

This certainly offers us a handle on the Correll cases.[15] There could be rational flaws at many stages in the process which leads up to the relevant experience: the visual system might put undue weight on stereotype-congruent priors, letting them "hack" the process of conditionalization. Or it could conditionalize normally but on priors that associate black men and crime to a degree that is out of all proportion with the evidence encountered. As a result, the "conclusion" of that process would be disproportionate to the legitimate data available to the system.

The second of these flaws in particular seems likely to pick out a very large proportion of instances of prejudicial bias, and this strategy will doubtless demarcate an important set of problematic cases. It has two limitations, however. The first is that it does not distinguish prejudicial bias as a distinctive phenomenon in its own right. Rather, it assimilates it to an existing category of epistemic flaw: irrationality. That may be the best we can do. Perhaps all that prejudicial bias amounts to is the subclass of formal bias which falls foul of some preexisting epistemic norm.

But a further concern we might have with this conclusion is not just that it fails to distinguish the intuitive set of cases we started with from many other cases of epistemically flawed belief, but that it may not even include every case that falls within the set of prejudicial bias. To see this, take the Correll case again, and suppose that a stereotype that does reflect the individual's evidence has been drawn on, proportionally, to inform the relevant experience. Such a case need not fall foul of evidentialist norms: the visual system is responding proportionately to the

[14] One could also adopt a version of this strategy the couched the evaluation in terms of reliability rather than rationality. The transformation would then be illegitimate because it is of an unreliable type.

[15] It is harder to see how this approach lets us say anything about the development of expertise in face perception. Claiming that this expertise is invariably out of proportion to the sample of faces the subject has encountered, for instance, seems unlikely to hold true in all relevant cases.

information at its disposal. In fact, to achieve the kind of intuitive neutrality we instinctively favor, what we want is for the visual system to substantially disregard some of the evidence it does have in favor of evidence it does not have. Equally, to take another candidate rational norm, prejudicial bias need involve no failure of *coherence* on the part of the individual or their visual system (BonJour 1985; Lehrer 1990). On the contrary, the problem seems to be rather than coherence is playing too large a role in informing the resultant state: coherence of visual experience with prior expectations itself strikes us as a flaw.

Can the problem be captured instead in terms of reliability? This brings us back to the observation we started with, that patterns of structural bias are an integral element in the normal functioning of the visual system. Given that fact, we are confronted with a nasty instance of the generality problem: how is the process responsible for prejudicial bias to be typed?[16] If we characterize it in formal terms it is likely to be of the same type as various other processes in visual perception that involve a similar use of priors in the interpretation of incoming information. That leaves us with no distinguishing marker between cases of prejudicial bias and normal functioning. How else can we type the process so as to avoid this result? Must we appeal to *content* as a way of distinguishing the relevant *processes?* This lands us back where we started, trying to find a distinctive structure, beyond content, that marks out instances of prejudicial bias within the visual system.

b. Neglect of available information, and cutting off enquiry

One particularly troubling feature of both the same-race face effect and the Correll cases is the way in which the subject has evidence at their disposal to which they fail to give adequate weight. If the subject in the Correll case only directed their gaze differently, they might gather counterevidence to the stereotype they rely on. Similarly, the individual whose face perception is skewed away from the detection of information from out-group faces has developed a visual practice that systematically neglects to pick up on readily available information in a subset of faces. When viewing an other-race face, there is further information directly within the subject's sight line to which they fail to respond. These practices of visual search are significantly negligent of *available information*. They are akin to an individual researching the flora of Madagascar who fails to open the book on their desk on precisely that subject, favoring instead to largely report their preexisting beliefs on the topic. Perhaps prejudicial visual bias involves a distinctive pattern of selection effect, resulting in the epistemically problematic neglect of available information.

Susanna Siegel (2013a) considers some of the epistemic impacts of selection effects in her discussion of visual experiences that are "anti-selected for uptake": experiences

[16] The generality problem is the problem of arriving at a principled means of "typing" the processes responsible for belief, in a manner that arrives a gives rise to a determinate assessment of their reliability (Conee and Feldman 1998). The reliability of the relevant process determines the degree of justification the belief enjoys depends.

which fail to be taken up for further processing, or to be drawn on in the formation of belief or execution of action. Siegel argues that the neglected experience retains evidential relevance and constitutes a defeater for the experience in question. In line with this, perhaps visual prejudicial bias involves a failure to uptake certain elements of an experience for processing. That neglected information can continue to serve as a defeater for the resulting experience.

Siegel's version of this response is couched in terms of the neglect of *experiences* which are already available to the individual in some form. As is stands, this will leave untouched cases in which the individual has *no such experience* at any level of visual processing. And it is not clear that in the problem cases considered above we are compelled to think that the relevant information from other-race faces, or from the experimental image of the target holding the object, is ever processed *at any level*. It needn't be the case that the subject has a fine-grained experience of an other-race face and then fails to "process" relevant information from it. They may only ever have a coarse-grained experience of it, that fails to deliver the information required for subtle emotion identification.

So we need to tweak this response to push the boundaries of the information counted as "available" out beyond the individual to include information that is *never* taken up to feature in an experience of any kind, but could easily be so. What seems culpable in these cases is the fact that subjects *don't* have such an experience, when relevant information was available that would have supported such an experience had they only attended to it. Perhaps the flaw then is that they cut off enquiry too soon. As Siegel writes, "sometimes it is epistemically appropriate to continue suspending judgement in the face of incoming information, rather than the settle on an answer to that question" (2013a:258). The information they do not yet have but could have had, had they approached their inquiry differently, undermines the good-standing of the belief they form on the basis of their visual experience.

This approach gives rise to the following candidate criterion for prejudicial bias: it involves a distinctive neglect of available information. The problem with face-expertise is that it causes us to systematically neglect certain sources of information. Similarly, the subjects in Correll's experiments have more information at their disposal, if they were only to move their eyes to the appropriate location. It is their pre-existing attitude to the question that their visual search is intended to settle that causes them to skip this step. This neglect is constitutive of prejudicial bias. This criterion explains in turn the various epistemic flaws prejudicial bias seems liable to give rise to: the resulting visual experiences are less reliable as a result, for instance, because available information was skipped over on the basis of that influence.

A standard of this kind would not be an *ad hoc* development for the visual case. Nathan Ballantyne argues for a broader norm for belief that allows that unpossessed evidence can serve as a defeater. Though Ballantyne's norm is couched in terms of belief formation, we could transpose it for our purposes onto the visual case:

unpossessed evidence could defeat the visual experience, rendering it, in effect, "irrational". Still, this approach fails to fit our purposes in two ways. In the first case, Ballantyne's arguments in favor of the norm are again grounded in a subject's possession of *evidence* of a defeater for the relevant belief.[17] The presence and problems of prejudicial visual bias seem consistent with the absence of any such evidence being in any sense available to the subject.[18] Secondly, this account over-predicts. We almost always have unpossessed evidence relevant to a given belief. Ballantyne is open to the possibility that this grounds a kind of broad skepticism about the epistemic status of our beliefs. But as we seek a norm capable of distinguishing cases of prejudicial bias, the global aspect of this skepticism is unhelpful. We want a feature specific to these cases.

This shortcoming helpfully points our way forwards, however: the difficulty posed by these cases is precisely that of pinning down when such neglect of information is illegitimate. We need to ask *why* we neglect information, and the worry is that the answer will not always negatively impact on the reliability of the visual system, nor constitute any deviation from the visual system's usual store of processing methods.

Consider cases of dramatic neglect of unexpected stimuli, as evidenced in instances of inattentional blindness. In these cases, observers routinely fail to spot an unexpected stimulus when engaged in a visual task that requires them to extract information of a certain kind from a display (when counting passes in a basketball-like game, you miss a man in a gorilla suit walking across the court (Chabris and Simons 2010), when counting touches between grey items you fail to notice a bright red cross moving across the field (Ward and Scholl 2015), when asked to identify nodules in images of lungs you miss a sketch of a gorilla in the background though your gaze is directed at it (Drew, Vo, and Wolfe 2013).) In these cases the visual system faces a pay-off between the information gained by devoting resources to spotting even the least expected of stimuli, and the information won by focusing instead on the task at hand and neglecting irrelevant stimuli. These do not seem like cases of prejudicial bias. In general, neglect of information is not necessarily irrational when your processing resources are limited. This criterion on its own cannot serve to distinguish prejudicial from systematic bias.

Perhaps prejudicial bias arises when differential quantities of attention are paid to a subset of stimuli, tracking a distinction that is irrelevant to the task at hand. Intuitively the problem with race-based facial expertise, the feature that makes it a potential case of *prejudicial* bias, is the *difference* in subjects' response to own-race and other-race faces. Similarly, the discrepant response to white and black men in the test stimuli is

[17] Ballantyne offers three distinct but related arguments that unpossessed evidence could serve as a defeater. All of them hinge either on the subject having evidence of the existence of a defeater, or reason to think that they have a defeater.

[18] Put simply, endorsing the principle that evidence of evidence is evidence (R. Feldman 2006 :223) doesn't get us far in these cases if the subject has no evidence of the relevant evidence.

what troubles us about the Correll cases. The subject's task is to determine whether the target is holding a gun or not, not to determine anything directly related to their race. So the adoption of different strategies when confronted with white and black individuals appears unwarranted.

The difficult question with which this strategy confronts us, is what makes a difference relevant or irrelevant to the task at hand? The explicit description of the Correll task makes no mention of race. But the individual's prior, we can hypothesize, attributes some higher likelihood of manifesting the relevant property, carrying a gun, to individuals of one race than another. Is it always illegitimate to draw on a prior association with a property not explicit in the task-description of the task? That seems too strong a restriction to place on the use of prior information. If I am engaged in a search for strawberries, I will do well to draw on a prior that strawberries are red, even though my search is not for strawberries under a description of them as red. To prohibit appeal to an association in virtue of its racialized content is to fall back to a characterization of prejudicial bias in contentful rather than formal terms.[19]

Perhaps prejudicial bias in these cases is marked by something else, for instance, the obvious relevance of the information that is neglected. In the Correll case, the task *just is* to identify the object the individual is holding. And yet eye-gaze tracking reveals that the subject, when confronted with a black individual, is looking not at the object the individual is carrying, but at their face. Disregarding highly relevant stimuli in this manner is akin to walking down the street trying to avoid stray debris, but looking at the street signs whilst doing so. Plausibly, any such irrational choice of information-gathering strategy would amount to an instance of prejudicial bias.

The worry for this approach is that Correll's characterization of this process supports a legitimizing interpretation of it. Correll suggests that the stereotypes are used to *supplement* the existing information. That is to say, reliance on that body of stored data is used to fill a shortfall of information in one area, leaving the individual free to gather additional data of some kind. This would be akin to a body of information about the likely location of debris in the street legitimately freeing an individual to gather additional data about the street signs. We need, then, some further reason to think that *this kind of information* is not a legitimate source, that the supplementation with the information from the stereotype does not give the subject good reason to stop their search. That is the possibility I consider next.

[19] Work in machine learning gives rise to cases that raise similarly difficult questions of when indirect tracking of racial properties via "proxy" properties (properties that correlate with, whilst not explicitly mentioning race) amounts to racial bias. For instance, algorithms drawn on in estimating risk when granting bail, that do not specify the race of the defendant as an input, can nonetheless end up with racially discrepant results on the basis of their reliance on properties that, in effect, serve as proxies for race (Corbett-Davies et al. 2017). I am indebted to Rob Long for discussion of these cases. For an argument that this kind of reliance on proxies could serve as a model of implicit bias in humans, see Johnson (ms.)

c. Arational, emotive attitudes

The mere facts of failing to gather more information, or of adopting strategies that distinguish between ostensibly irrelevant subcategories are not sufficient to distinguish prejudicial bias. Another possible route to doing so appeals to facts about the *kind* of state responsible for the influence. On this approach, prejudicial bias involves the neglect of information driven not by proportionate priors, but by arational affective attitudes for instance.

Perhaps the problem in the Correll cases lies in the visual system's appeal to states such as *stereotypes*. I will take a stereotype to be a cluster of information associated with a particular social group, information which can take a variety of forms including propositional beliefs and affective attitudes.[20] For the purposes of this approach, should we assume that stereotypes have a distinctively negative epistemic valence? The existing literature is split on this question, sometimes treating the category as epistemically neutral, and sometimes assuming that it is definition of a stereotype that it is epistemically flawed.[21] We face a dilemma whichever way we go on this question. If we use "stereotype" to indicate a distinctively *inaccurate* collection of information, then the problem with its involvement in these cases lies not in its classification as a stereotype, but in its inaccuracy. If, on the other hand we draw the class of stereotypes such that it includes accurate instances, then we are owed an explanation of why such states cannot legitimately serve as stores of prior information.

What we have in mind by appealing to stereotypes to mark out cases of prejudicial bias may be not their inaccuracy, but the possibility that they involve not only belief-like attitudes which could be true or false, or otherwise proportionate to the evidence, but in addition arrational affective attitudes that drive the accumulation of information in these cases. Affective attitudes may also be in play in the case of face-perception. Expertise with own-race faces is partly driven by the close emotional attachments infants and children develop with their caregivers. We can set aside the label of stereotype and look instead just at this possibility: that prejudicial bias is distinguished by the influence of arational affective attitudes on cognitive processing. In virtue of their arationality, such affective attitudes are liable to "hack" the processes by which new and old information are integrated, thereby resulting in beliefs or experiences that fail to be proportionate to the evidence, but that are instead skewed to support the existing emotive attitudes of the individual.

Pinning the blame on the drive of an affective attitude requires us to point in turn to a principled reason why affect in particular is an invariably illegitimate influence.

[20] This is intended to be as neutral as possible a definition of a contested term that enjoys a wide range of uses in both colloquial conversation and academic work. See Madva and Brownstein (2016) for a helpful taxonomy of its usage.

[21] See the work of Lee Jussim and his collaborators for arguments that stereotypes are frequently accurate, in the sense that they are statistically borne out (Jussim et al. 2009; Madon et al. 1998). Others take their inaccuracy to be a definitional feature of them (For a perspective from within social psychology, see Jost and Banaji 1994; For a philosophical perspective see Blum 2004).

Affective states too can encode previously encountered information. Consider, for instance, the fear you feel entering notoriously shark-infested water. That fear is an affective attitude, but it can also be proportionate to the evidence you have of the risk of shark attack.[22] And it can, moreover, skew results in ways that seem epistemically beneficial: you are now *more likely* to detect sharks because your fear makes you appropriately sensitive to possible stimuli that you would not normally spot, such as fin-like protrusions, or gray shapes beneath the surface of the water. There is room for these attitudes to exert a legitimate influence on the direction of attention despite their affective form. If the involvement of affective attitudes alone were sufficient for prejudicial bias, the shark case would have to also be counted as an instance.

All of these potential criteria will undoubtedly circumscribe some set of cases, some instances of which will be particularly problematic, and embody prejudicial bias. But none seem to catch *only* classes of prejudicial bias, nor do any of them seem capable of catching *all* cases of prejudicial bias. For in every case we can imagine an instance of prejudicial bias which *does* conform to the standard in question, but still has prejudicial upshots. And similarly we can imagine innocuous cases of merely formal bias which are caught by the relevant criterion. In light of this impasse I turn in the next section to consider *why* these standards are failing to capture the relevant phenomenon.

3. A skill-based account of visual prejudicial bias

Why are these epistemic norms failing to locate a problem with these particular cases? One reason is that standard forms of epistemic evaluation focus on transitions between pieces of information, or the truth ratio of processes that govern such transitions. In this way, they apply to processes that operate over a previously fixed body of information. We naturally think of perception as though it were such a process: the immediate environment is fixed, and thereby fixes the information available to the open eye as it views it. But that natural thought is misleading. Given the processing limitations on visual perception, it has to be *selective* even within a fixed environment, and to *interpret* the data it receives. As a result, a significant part of the perceptual process involves a series of decisions about which available stimuli to focus on. What we need in order to capture the cases of visual prejudicial bias described above, is an evaluation of the processes by which certain pieces of information, and certain interpretations of them, are prioritized over others.[23]

But it is hard to perform such an evaluation in purely formal terms, because these choices involve practical pay-offs. Expertise and the exploitation of information in

[22] See Wells and Matthews(2014) for a full discuss of the important role emotion plays in directing perceptual attention. For a discussion of the capacity of emotions to constitute a proportionate or rational response, see Turski (1994).

[23] We are, in effect, in need of a set of norms of *inquiry*. It is unclear what the relationship is between norms of inquiry and epistemic norms more broadly. Whilst it might seem natural to treat the former as a subset of the latter, see Friedman (m.s.) for an argument that norms of inquiry, what she terms "zetetic norms", may be systematically in conflict with certain epistemic norms.

one area comes at the cost of maximal exploitation and expertise elsewhere. Reliance on malleable priors inevitably give rise to a dilemma, between priors that adapt to a local context, and flexibility within a broader, global context. What settles the point at which the visual system best balances local specialization and global flexibility depends not on something intrinsic to the visual system, or to the body of data within its existing compass, but on whatever pay-off best serves the organism's purposes in a current context. In some situations, it may be more valuable to us to win greater accuracy in a local context at the price of global neutrality: for instance, once expert at detecting nodules in images of lungs, it is no longer worth our while to keep a look out for highly unexpected stimuli in that context. But in other cases, suboptimal performance at a local level (manifested, for instance, by the slower identification of repeated stimuli), might be worth a gain in neutrality at a global level, (such as an enhanced capacity to detect novel features of a changing environment). In the case of visual prejudice, valuable expertise with local race-based regularities: (faces, or other local associations), loses us global neutrality, a highly valued epistemic good in the context of our reasoning about other people.

The instances of prejudicial bias within the visual system described above consist in the preferential selection of information. In the Correll case, the visual systems prefers to rely on old information, rather than undertaking a more active inquiry into its current environment. In the case of face perception, the development of expertise allows for maximal extraction of information from own-race faces, to the sacrifice of possible information from other-race faces. Formal descriptions of the method by which information, once acquired, is integrated, or epistemic norms that exclusively evaluate that process, are poorly placed to identify or evaluate bias at this prior stage, the stage at which information is acquired.

In fact, this sort of decision *cannot* be evaluated in purely formal terms that abstract away from the content of the relevant information, because the relative value of that content in a given context *is* the basis on which the decision is made. To evaluate that decision we need to know how well it serves the various goals of the organism. Nor can we simply plug in a generic epistemic goal such as accuracy, or truth.[24] It isn't that one or other decision gets us more or less accuracy so much as *different* accuracy: specialization within *this face space* lets you access accurate information of a certain kind. Specialization within a different face space lets you access information of a different kind. Reliance on priors to the neglect of certain stimuli frees you up to gather information elsewhere. Norms that dictate an attentional strategy relative to set of facts one seeks to learn do not go far enough. Even what set of facts one should seek to know depends in turn on the goals the inquiry serves.

Successful visual perception does not just involve the acquisition of information per se, but the acquisition of the information which best positions us to achieve our goals. Both the goal-directed aspect of visual perception, and its capacity to manifest

[24] For the claim that truth is the primary epistemic goal, see Hess (2010). For an instance of an accuracy-first approach see Joyce (1998).

expertise, assimilate it to a kind of *skill*. Thinking of it in these terms offers us a framework within which to analyze the impact of its reliance on various biases on its capacity to achieve competing epistemic goals.

Skilled activity is the capacity to achieve a goal through practice within a given environment. We cannot evaluate a skill except relative to a goal, and environment. There is no absolute standard for assessing tennis skill outside of the parameters provided by conventions that govern the game. Moreover, what constitutes tennis skill depends on features of the environment: a competent grass-court player may struggle on a clay surface.

I propose a schema for the evaluation of skilled activity of the following form:

Skill: A individual *S*, is skilled at activity Φ relative to a goal *g*, environment *e,* and over timeframe *t* iff their practice of Φ in environment *e* within timeframe *t* is likely to position them to achieve goal *g*.

$([[\text{“is skilled at } \Phi\text{-ing”}]]^{c}{}_{<g,tf,e>} = \{x \mid x \text{ is skilled at } \Phi\text{-ing relative to a goal } g, \text{ in environment } e, \text{ over time-frame } t.)$

This general schema applies to the case of perceptual skill. There is no absolute standard of perceptual skill outside of the epistemic and practical goals we have at any one particular moment. An individual's skill at seeing can only be evaluated relative to a goal, within a given environment and across some set timeframe. The specification of the timeframe is closely tied to the individuation of the environment: how we delineate the time-frame will determine the quantity of change in the environment to which the visual system must adapt. For instance, if the context is construed to include a timeframe during which it grows dark and then light again, a very different perceptual profile will count as skilled than if the timeframe only includes daylight periods.

For our purposes, a crucial upshot of this schema for the evaluation of skill is that it will deliver multiple competing evaluations relative to multiple different goals which the individual may have. What constitutes skilled seeing relative to one goal may not constitute skilled seeing relative to another. In fact, the goals that perception serves can frequently come apart. Construed very broadly, they will include both gathering information and effective action.[25] Even within the subset of an organisms' epistemic goals, there will be divergent ends such as accurately representing repeated local stimuli, or retaining a sensitivity to unexpected novel items. We can derive competing evaluations of their perceptual skill relative to these different goals. There will be no single verdict on an individual's perceptual skill without an ordering on goals that lets

[25] For discussion of object-directed action as a function of the visual system, and the interaction of that goal with the goal of accurate representation of the environment see Goodale and Humphrey (1998).

us integrate these competing evaluations. Ultimately, verdicts relative to one or other goal may remain more informative than any such synthesis.

How does this help us identify bias in the cases above? Just as evaluation of skill has to take place relative to a goal, so the identification of prejudicial bias can only happen relative to a goal. Prejudicial bias limits our capacity to achieve certain goals: more specifically, intuitively problematic prejudicial bias inhibits our ability to accomplish particular set of epistemic goals that relate to the neutral treatment of others in our environment. But consistent with the recognition of competing goals above, what constitutes a prejudicial bias relative to that goal may facilitate the achievement of other goals.[26]

We can now define prejudicial bias: *an instance of formal bias is an instance of prejudicial bias when its effect is to impede demographically-neutral epistemic treatment of other individuals.*

We do not treat others in an epistemically neutral fashion when their demographic status plays a significant role in determining how we acquire information about them. Demographically-neutral epistemic treatment of others is inconsistent with appeal to base rates that encode social categories. Similarly, it prohibits the adoption of strategies for the extraction of information that vary depending on the social profile of an individual. This leaves much unspecified: what impediment is sufficient to warrant the classification, and what constitutes "neutral treatment" will vary depending on context.[27] But the definition at least gives us a handle on the way in which prejudicial bias is identified relative to a goal. We have a multitude of reasons for prizing the neutral epistemic treatment of other individuals. Some of these are ethical: doing so promotes the right treatment of others. Some of them are distinctively epistemic: we value neutrality because it preserves our capacity to access information across different contexts, information that is particularly valuable to us when it concerns other individuals.[28] Our criterion for distinguishing prejudicial bias is not formal: to apply it we have to attend to the contents of the relevant states. But prejudicial bias is still an epistemically significant category, because the goal in question is an epistemic one. We value the capacity to think about others in ways that

[26] See Gendler (2011) for discussion of the epistemic dilemma that race-based priors give rise to, as both encoding them, and failing to encode them, appear to carry epistemic costs.
[27] It might look like this over-predicts instances of prejudicial bias. It identifies prejudicial bias, for instance, when a doctor uses information about demographic categories to predict risk. I accept that these cases *are* instances of prejudicial bias, but instances whose utility relative to other practical and epistemic goals outweighs the disutility involved in the loss of neutrality.
[28]My claim here is that prejudicial bias is a category defined relative to an epistemic goal that is liable to have further knock-on moral and epistemic consequences. Moss (2018) proposes another model for how the moral and the epistemic interact in the case of racial statistical belief, according to which the moral stakes of forming beliefs about an individual encroach on epistemic standards. In particular we are subject to a moral obligation to keep in mind the possibility that the individual is an exception to the relevant statistic. Keeping that in mind has the result that we fail to *know* the inference from the statistic as applied to the individual. Basu (ms., ms.b) argues for a kind of moral encroachment in cases of racist belief, and offers a rich discussion of the epistemic and ethical costs of such beliefs.

disregard their demographic status. It is against this goal that the category of prejudicial bias is defined.

Take the case of face perception. The specialization of one's capacities for facial recognition within a particular area of face space is likely to serve us well relative to the goal of reading and remembering the faces of those in our immediate familial or social circle. But we also value highly the ability to accurately read the faces of other individuals who fall outside of that group. This latter goal falls under the broader aegis of the privileged goal described above, it constitutes demographically-neutral epistemic treatment of others. So this specialization, though it constitutes a kind of skilled expertise on one axis of evaluation, is an instance of prejudicial bias. Similarly, a subject's reliance on stereotype-consistent priors may facilitate speeded recognition of a certain range of objects within a stable environment, but it is an impediment to other goals which again fall into the set against which prejudicial bias is defined, namely, the goal of retaining a kind of racial neutrality when perceiving individuals.
Thinking of perception as a skill has other helpful upshots in our endeavor to understand the nature of perceptual bias, both formal and prejudicial. In the first place, doing so moves us towards an understanding of visual perception as an *active* process that responds dynamically to environmental challenges. In practicing a skill like tennis, or cookery, we build through repetition a set of mental states capable of appropriately guiding the activity in question.[29] Similarly, by repeatedly seeing a particular environment, we build priors that optimally guide our perception of that environment.

This in turn directs our attention to the role of the context in which the activity is practiced in honing the resultant skill. No matter how gifted an individual practitioner, her skills are inevitably limited by the environment she finds herself in. Practicing running on one particular surface hones our skills at running on that particular surface. An excessively narrow training environment may limit our capacity to flexibly adapt to a new surface. A bias need not be rooted in an individual performance, but in the field the game is played on, or even the field the player has consistently practiced on.

Perceptual skill, too, recapitulates the learning environment. One way of avoiding prejudicial visual bias is to ensure the individual "practices" on an appropriately varied set of samples. But his opportunities to do so may depend significantly on facts about social organization: practices of segregation, implicit or explicit, limit the learning sample and with it the flexibility of the resulting skill. In doing so, they give rise to manifestations of visual prejudicial bias.[30]

---

[29] This meshes with Stanley and Williamson's (2016) definition of skill as "a disposition to form knowledge states appropriate for guiding" the activity in question.
[30] See Munton (forthcoming) for an in-depth argument that social structures can cap perceptual skill, in particular via the sensitivity of visual priors to regularities secured by structural injustice.

4. Conclusion

The visual system does not simply respond to a determinate set of information. It *selects* information on the basis of past exposure and present motivation. That selective process opens the door to varieties of prejudicial bias that we struggle to capture in the terms of epistemic analyses designed primarily for evaluating transitions *within* a fixed body of information. Recognizing the ways in which the visual system is an active, selective process should encourage us to think of it as kind of skill, and to adopt an evaluative framework appropriate to its status as such.

Doing so opens the way for us to identify prejudicial bias within a neutrally biased system. That identification must take place relative to a goal the individual tries to achieve. We have a standing goal, of preserving a kind of epistemic neutrality towards others regardless of their demographic properties, which certain cases of bias fall foul of. These are cases of prejudicial bias. Although prejudicial bias is defined in relation to an epistemic goal, the way in which it can arise in the *acquisition* of information (rather than in aberrant transitions *between* pieces of information), renders standard tools of epistemic evaluation poorly placed to identify it within a structurally biased system such as visual perception.

Works Cited

Basu, Rima. ms. "Locating the Wrong in Racist Beliefs."

———. ms. "The Moral Stakes of Racist Beliefs."

Blum, Lawrence. 2004. "Stereotypes And Stereotyping: A Moral Analysis." *Philosophical Papers* 33 (3):251–289.

BonJour, Laurence. 1985. *The Structure of Empirical Knowledge*. Harvard University Press.

Brownstein, Michael, and Jennifer Saul. 2016. *Implicit Bias and Philosophy, Volumes 1 and 2: Metaphysics and Epistemology; Moral Responsibility, Structural Injustice, and Ethics*. Oxford, New York: Oxford University Press.

Catz, Or, Michal Kampf, Israel Nachson, and Harvey Babkoff. 2009. "From Theory to Implementation: Building a Multidimensional Space for Face Recognition." *Acta Psychologica* 131 (2):143–52. https://doi.org/10.1016/j.actpsy.2009.03.010.

Chabris, C.F., and D.J. Simons. 2010. *The Invisible Gorilla: How Our Intuitions Deceive Us*. New York: Random House.

Clifford, Colin WG, Michael A Webster, Garrett B Stanley, Alan A Stocker, Adam Kohn, Tatyana O Sharpee, and Odelia Schwartz. 2007. "Visual Adaptation: Neural, Psychological and Computational Aspects." *Vision Research* 47 (25):3125–3131.

Conee, Earl, and Richard Feldman. 1998. "The Generality Problem for Reliabilism." *Philosophical Studies* 89 (1):1–29.

Connee, Earl, and Richard Feldman. 2004. *Evidentialism: Essays in Epistemology*. Oxford University Press.

Corbett-Davies, Sam, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. "Algorithmic Decision Making and the Cost of Fairness." *CoRR* abs/1701.08230. http://arxiv.org/abs/1701.08230.

Correll, J., B. Park, C.M. Judd, and B. and Wittenbringk. 2007. "The Influence of Stereotypes on Decision to Shoot." *European Journal of Social Psychology* 37 (1102–1117).

Correll, Joshua, Matthew Crawford, Bernd Wittenbrink, and Melody Sadler. 2015. "Stereotypic Vision: How Stereotypes Disambiguate Visual Stimuli." *Journal of Personality and Social Psychology* 108 (2):219–233.

Cushion, Stephen, Kerry Moore, and John Jewell. 2011. "Media Representations of Black Young Men and Boys: Report of the REACH Media Monitoring Project." Department for Communities and Local Government, London.

Davenport, J.L., and M.C. Potter. 2004. "Scene Consistency in Object and Background Perception." *Psychological Science* 15:559–564.

Drew, T., M. Vo, and J. Wolfe. 2013. "The Invisible Gorilla Strikes Again: Sustained Inattentional Blindness in Expert Observers." *Pyschological Science* 24:1848–1853.

Eberhardt, Jennifer L., Phillip Atiba Goff, Valerie J. Purdie, and Paul G. Davies. 2004. "Seeing Black: Race, Crime, and Visual Processing." *Journal of Personality and Social Psychology* 87 (6):876–93. https://doi.org/10.1037/0022-3514.87.6.876.

Fantl, Jeremy, and Matthew McGrath. 2002. "Evidence, Pragmatics, and Justification." *Philosophical Review* 111 (1):67–94.

Feldman, J. 2014. "Bayesian Models of Perceptual Organization In." In *Oxford Handbook of Perceptual Organization*, edited by J. Wagemans.

Feldman, Richard. 2006. "Epistemological Puzzles About Disagreement." In *Epistemology Futures*, edited by Stephen Hetherington, 216–236. Oxford University Press.

Friedman, Jane. m.s. "The Epistemic and the Zetetic."

Gauthier, Isabel, Michael Tarr, and Daniel Bub. 2009. *Perceptual Expertise: Bridging Brain and Behavior*. Oxford University Press.

Ge, Liezhong, Hongchuan Zhang, Zhe Wang, Paul C Quinn, Olivier Pascalis, David Kelly, Alan Slater, Jie Tian, and Kang Lee. 2009. "Two Faces of the Other-Race Effect: Recognition and Categorisation of Caucasian and Chinese Faces." *Perception* 38 (8):1199–1210. https://doi.org/10.1068/p6136.

Gendler, T. 2011. "On the epistemic costs of implicit bias." *Philosophical Studies* 156 (1):33–63.

Goldman, Alvin. 1979. ""What Is Justified Belief?"." In *Justification and Knowledge*, edited by George Pappas, 1–25. Boston: D. Reidel.

Goodale, Melvyn A, and G Keith Humphrey. 1998. "The Objects of Action and Perception." *Cognition* 67 (1):181–207. https://doi.org/10.1016/S0010-0277(98)00017-1.

Greenwald, Anthony G, and Linda Hamilton Krieger. 2006. "Implicit Bias: Scientific Foundations." *California Law Review* 94 (4):945–967.

Harel, Assaf, Dwight Kravitz, and Chris I. Baker. 2013. "Beyond Perceptual Expertise: Revisiting the Neural Substrates of Expert Object Recognition." *Frontiers in Human Neuroscience* 7. https://doi.org/10.3389/fnhum.2013.00885.

Hawthorne, John. 2003. *Knowledge and Lotteries*. Oxford University Press.

Hess, Markus Patrick. 2010. *Is Truth the Primary Epistemic Goal?* Vol. 17. Walter de Gruyter.

Jaquet, Emma, and Gillian Rhodes. 2008. "Face Aftereffects Indicate Dissociable, but Not Distinct, Coding of Male and Female Faces." *Journal of Experimental Psychology. Human Perception and Performance* 34 (1):101–12. https://doi.org/10.1037/0096-1523.34.1.101.

Jaquet, Emma, Gillian Rhodes, and William G. Hayward. 2007. "Opposite Aftereffects for Chinese and Caucasian Faces Are Selective for Social Category Information and Not Just Physical Face Differences." *The Quarterly Journal of Experimental Psychology* 60 (11):1457–67. https://doi.org/10.1080/17470210701467870.

———. 2008. "Race-Contingent Aftereffects Suggest Distinct Perceptual Norms for Different Race Faces." *Visual Cognition* 16 (6):734–53. https://doi.org/10.1080/13506280701350647.

Johnson, GM. ms. "Algorithmic Bias: On the Implicit Biases of Social Technology."

Jost, John T, and Mahzarin R Banaji. 1994. "The Role of Stereotyping in System-Justification and the Production of False Consciousness." *British Journal of Social Psychology* 33 (1):1–27.

Joyce, James M. 1998. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science* 65 (4):575–603.

Jussim, L., T. Cain, J. Crawford, K. Harber, and F. Cohen. 2009. "The Unbearable Accuracy of Stereotypes." In *Handbook of Prejudice, Stereotyping and Discrimination. Erlbaum*, edited by T. Nelson, 199–227.

Kelly, David J, Paul C Quinn, Alan M Slater, Kang Lee, Liezhong Ge, and Olivier Pascalis. 2007. "The Other-Race Effect Develops during Infancy: Evidence of Perceptual Narrowing." *Psychological Science* 18 (12):1084–1089.

Kersten, Daniel, and Alan Yuille. 2003. "Bayesian Models of Object Perception." *Current Opinion in Neurobiology* 13 (2):150–158.

Kornblith, Hilary. 2002. *Knowledge and Its Place in Nature*. Oxford University Press.

Lehrer, Keith. 1990. *Theory of Knowledge*. Routledge.

Levin, Daniel. 1996. *Classifying Faces by Race: The Structure of Face Categories*. Vol. 22. https://doi.org/10.1037/0278-7393.22.6.1364.

———. 2000. "Race as a Visual Feature: Using Visual Search and Perceptual Discrimination Tasks to Understand Face Categories and the Cross-Race Recognition Deficit." *Journal of Experimental Psychology: General* 129 (4):559–74. https://doi.org/10.1037/0096-3445.129.4.559.

Little, Anthony C, Lisa M DeBruine, and Benedict C Jones. 2005. "Sex-Contingent Face after-Effects Suggest Distinct Neural Populations Code Male and Female Faces." *Proceedings of the Royal Society B: Biological Sciences* 272 (1578):2283–87. https://doi.org/10.1098/rspb.2005.3220.

Madon, S.J., L. Jussim, S. Keiper, J. Eccles, A. Smith, and P. Palumbo. 1998. "The Accuracy and Power of Sex, Social Class and Ethnic Stereotypes: Naturalistic Studies in Person Perception." *Personality and Social Psychology Bulletin* 24:1304–1318.

Madva, Alex, and Michael Brownstein. 2016. "Stereotypes, Prejudice, and the Taxonomy of the Implicit Social Mind1." *Noûs* 50 (4).

Meissner, Christian A, and John C Brigham. 2001. "Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review." *Pyschology, Public Policy and Law* 7 (1):3–35.

Moss, Sarah. 2018. *Probabilistic Knowledge*. Oxford University Press.

Munton, Jessie. forthcoming. "Perceptual Skill and Social Structure." *Philosophy and Phenomenological Research*.

National Center for Education Statistics. 2016. "Yale University." 2016. https://nces.ed.gov/ipeds/datacenter/Facsimile.aspx?unitid=acaeabb2b4af.

Palmer, S.E. 1975. "The Effects of Contextual Scenes on the Identification of Objects." *Memory and Congition* 3:519–526.

Payne, B. Keith. 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. Keith, Yujiro Shimizu, and Larry L. Jacoby. 2005. "Mental Control and Visual Illusions: Toward Explaining Race-Biased Weapon Misidentifications." *Journal of Experimental Social Psychology* 41 (1):36–47. https://doi.org/10.1016/j.jesp.2004.05.001.

Rescorla, M. 2015. "Bayesian Perceptual Psychology." In *The Oxford Handbook of the Philosophy of Perception. OUP*, edited by M. Matthen.

Rhodes, Gillian, and David A. Leopold. 2011. "Adaptive Norm-Based Coding of Face Identity," July. https://doi.org/10.1093/oxfordhb/9780199559053.013.0014.

Rhodes, Gillian, Vance Locke, Louise Ewing, and Emma Evangelista. 2009. "Race Coding and the Other-Race Effect in Face Recognition." *Perception* 38 (2):232–41. https://doi.org/10.1068/p6110.

Scholl, B.J. 2005. "Innateness and (Bayesian) Visual Perception: Reconciling Nativism and Development." In *The Structure of the Innate Mind*, edited by P. Carruthers, S. Laurence, and S. Stich. Cambridge: Cambridge University Press.

Siegel, S. 2013a. "Can Selection Effects on Experience Influence Its Rational Role?" In *Oxford Studies in Epistemology*, edited by T. Gendler, 4:240. Oxford.

———. 2013b. "The Epistemic Impact of the Etiology of Experience." *Philosophical Studies* 162 (3):697–722.

Siegel, Susanna. 2010. *The Contents of Visual Experience*. Oxford University Press USA.

———. 2016. *The Rationality of Perception*.

Stanley, Jason. 2005. *Knowledge and Practical Interests*. Oxford University Press.

Stanley, Jason, and Timothy Williamson. 2016. "Skill." *Noûs*.

Summerfield, C., and T. Egner. 2009. "Expectation (and Attention) in Visual Cognition." *Trends in Cognitive Science* 13 (9):403–409.

Turski, W. George. 1994. *Toward a Rationality of Emotions: An Essay in the Philosophy of Mind*. Athens: Ohio University Press.

Valentine, Tim. 1991. "A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition." *The Quarterly Journal of Experimental Psychology* 43 (2):161–204.

Valentine, Tim, Michael B Lewis, and Peter J Hills. 2016. "Face-Space: A Unifying Concept in Face Recognition Research." *The Quarterly Journal of Experimental Psychology* 69 (10):1996–2019.

Ward, E.J., and B.J. Scholl. 2015. "Inattentional Blindness Reflects Limitations on Perception, Not Memory: Evidence from Repeated Failures of Awareness." *Psychonomic Bulletin & Review* 23:797–802.

Weatherson, Brian. 2012. "Knowledge, Bets, and Interests." In *Knowledge Ascriptions*, edited by Jessica Brown and Mikkel Gerken, 75–103. Oxford University Press.

Wells, Adrian, and Gerald Matthews. 2014. *Attention and Emotion (Classic Edition): A Clinical Perspective*. Psychology Press.

Wolfe, Jeremy M., and Todd S. Horowitz. 2017. "Five Factors That Guide Attention in Visual Search" 1 (March):0058.

Yale Office of Institutional Research. 2016. "Yale University Ladder Faculty Diversity 2005-2016." Tableau Software. 2016. https://public.tableau.com/views/YaleUniversityLadderFacultyDiversity2005-2015/Rank?:embed=y&:showVizHome=no&:host_url=http%3A%2F%2Fpublic.tableau.com%2F&:tabs=yes&:toolbar=yes&:animate_transition=yes&:display_static_image=no&:display_spinner=no&:display_overlay=yes&:display_count=yes&:loadOrderID=0.