Bootstrap Hell: Perceptual Racial Biases in a Predictive Processing Framework

Zachariah A. Neemeh (zaneemeh@memphis.edu)
Department of Philosophy, University of Memphis
Memphis, TN 38152 USA

Abstract

Predictive processing is transforming our understanding of the brain in the 21st century. Whereas in the 20th century we understood the brain as a passive organ taking in information from the world, today we are beginning to reconceptualize it as an organ that actively creates and tests hypotheses about its world. The promised revolution of predictive processing extends beyond cognitive neuroscience, however, and is beginning to make waves in the philosophy of perception. Andy Clark has written that predictive processing creates a "bootstrap heaven," enabling the brain to develop complex models of the world from limited data. I argue that the same principles also create a "bootstrap hell," wherein prejudice biases inherent in our inegalitarian societies result in permanent perceptual modifications. These modifications are unamenable to conventional implicit bias training. The deep embeddedness of prejudice biases in perceptual experience makes any proposal to eliminate prejudice biases by mere "understanding" insufficient.

Keywords: philosophy of cognitive science; predictive coding; predictive processing; racial bias

Introduction

Predictive processing, or predictive coding,[1] is transforming our knowledge of perception (Knill & Richards, 1996; Rao & Ballard, 1999), the brain (Friston, 2018; Hohwy, 2013; Knill & Pouget, 2004), and embodied cognition (Allen & Friston, 2018; Clark, 2016; Gallagher & Allen, 2018; Seth, 2015). Predictive processing is a hierarchical implementation of empirical Bayes, wherein the cognitive system creates generative models of the world and tests its hypotheses against incoming data. It is hierarchical insofar as the predictions at one level are tested against incoming signals from the lower level. The resulting prediction error, the difference between the expectation and the incoming data, is used to recalibrate the model in a process of prediction error minimization. Predictions may be mediated by pyramidal cells across the neocortex (Bastos et al., 2012; Hawkins & Ahmad, 2016; Shipp et al., 2013). Andy Clark has characterized predictive processing as creating a "bootstrap heaven" (2016, p. 19), enabling the brain to develop complex models of the world from limited data. This enables us to extract patterns from ambiguous signals and establish hypotheses about how the world works.

The training signals that we get from the world are, however, biased in all the same unsightly ways that our societies are biased: by race, gender, socioeconomic status, nationality, and sexual orientation. The problem is more than a mere sampling bias. Our societies are replete with prejudice biases that shape the ways we think, act, and perceive. Indeed, a similar problem arises in machine learning applications when they are inadvertently trained on socially biased data (Avery, 2019; N. T. Lee, 2018). The basic principle in operation here is "garbage in, garbage out": a predictive system that is trained on socially biased data will be systematically biased in those same ways.

Unfortunately, we are unwittingly trained on this prejudiced data from our earliest years. As predictive systems, we bootstrap upwards into more complex cognitive processes while being fed prejudiced data, spiraling us into a "bootstrap hell." This has repercussions for everything from higher-order cognitive processes down to basic perceptual processes. Perceptual racial biases include perceiving greater diversity and nuance in the faces of racial ingroup faces (the cross-race effect; Malpass & Kravitz, 1969), misperceiving actions of racial outgroup members as hostile (Pietraszewski et al., 2014), and empathetically perceiving emotions in racial ingroup (but not outgroup) faces (Xu et al., 2009), among other phenomena. They are particularly worrying due to their recalcitrance to conscious control or implicit bias training. We may be able to veto a prejudiced thought (but see Kelly & Roedder, 2008), but we cannot simply modify our perceptual experience at will. Recalcitrant predictions such as this are "hyperpriors" and are unamenable to rapid, conscious adjustment.

I begin with an overview of predictive processing. I explain that the same principles that allow us to bootstrap our way into full cognition also allow for biases to develop. These biases include perceptual racial biases, which are visual and affective rather than cognitive. I explain how sampling biases in infancy and emotion perception contribute to perceptual racial biases (although many other factors certainly play a role). Finally, I hypothesize that traditional implicit bias training may not be enough to disentangle the web of hypotheses that contribute to perceptual racial bias.

[1] The Bayesian paradigm goes under various guises, including predictive coding, predictive processing, and predictive engagement. Although there are differences in these approaches, I use the term "predictive processing" in an inclusive sense.

Bootstrap Heaven and Bootstrap Hell

From Helmholtz to the Bayesian Brain and Beyond

The idea that the brain produces hypotheses about the causes of its perceptual inputs is not a new one. Hermann von Helmholtz (1867/1925, 1878/1995) in the 19th century and Richard L. Gregory (1966/1997) in the 20th century have been the greatest champions of just such a view. According to Helmholtz (1867/1925), perception results from unconscious inferences. Helmholtz (1878/1995) conceived of these unconscious inferences in terms of syllogistic (i.e., Aristotelian) logic. Today, we understand unconscious inference to be largely Bayesian rather than syllogistic (Gregory, 2006). Originally applied to vision (Rao & Ballard, 1999; Srinivasan et al., 1982), predictive processing has made a rapid and stellar rise, becoming a dominant paradigm in cognitive neuroscience (Friston, 2018) and recently making waves in philosophy (Clark, 2016; Hohwy, 2013). While in the 20th century we saw the brain as a passive organ integrating external signals from the bottom-up—such as in Marr's (1982/2010) classical work on vision—in the 21st century we now see the brain as taking an active role in constructing its world from the top-down—"hallucinating" a world, in Metzinger's (2003) poignant phrase.

At the most schematic level, predictive processing is a looping series of precision-weighted predictions arranged hierarchically. The brain implements an approximation of Bayesian statistics for perception, action, reasoning, and other cognitive processes. A Bayesian brain would need vast computational resources in order to calculate probabilities fully. Fortunately, the brain does not need to calculate explicit probabilities in order to approximate Bayesian statistics. More modestly, it may be a Bayesian sampler, implementing Markov chain Monte Carlo, for example (Sanborn & Chater, 2016). Bayesian statistics allows for a continual updating of prior probabilities, also referred to as 'priors,' 'hypotheses,' or 'beliefs.' Priors are tested against an incoming signal. The difference between the two results in "surprise," generating a predictive error signal. This allows the priors to be reweighted, resulting in a more accurate prior. This process is hierarchical, and at each level in the hierarchy, the process models the level below it. Predictions are precision-weighted. When a prediction mismatches the input, a prediction error signal is propagated upwards and triggers the reweighting of priors. Empirical evidence that the brain is structured in just such a hierarchical, predictive form is preliminary but mounting (Clark, 2013, 2016; Friston, 2008).

Bootstrap Heaven

Predictive processing allows for more than an understanding of neural circuits; it also allows us to understand how human cognition develops in an impoverished environment. Predictive processing creates a "bootstrap heaven," allowing the brain to bootstrap itself up into more complex predictions given relatively limited and basic inputs (Clark, 2016, p. 19). Essentially, we are fed training signals, and our brains develop predictive models of their inputs.

A predictive system is well-equipped to deal with the poverty of the stimulus. For example, while no limited number of sentences may be sufficient to deduce the full rules of grammar (Chomsky, 1959), they may well be sufficient to create predictive models of grammar through grammatical inference (Clark, 2016). The same bootstrapping principle operates in perception. The brain develops hypotheses about its sensory inputs. These inputs are ambiguous, as formulated in the inverse projection problem: projecting the cause of a stimulus from a sensory pattern always leaves ambiguity. However, perceptual hypotheses operate on multiple spatiotemporal scales, allowing for the salience of context to guide the interpretation of a signal (Pizlo, 2001). Contextual factors allow for the minimization of ambiguity in inverse projection. Predictions become more granular at lower levels of the hierarchy. A coherent set of mutually-supporting hypotheses at multiple spatiotemporal scales allows for the top-down influence of context to drive the prediction of an incoming signal.

Bootstrap Hell

Generally, aberrant predictions normalize out due to the constant and hierarchical updating of priors through the propagation of prediction error signals. Our first ever sight of a banana—let us say that it is yellow—will result in our model of bananas being unduly weighted towards yellowness. Only after many trials in perceiving bananas will our model of them include a range of colors from yellow to black. Our model is gradually reconfigured as we encounter novel situations and objects engendering prediction errors, such as a black or spotted banana. The initially aberrant model was based on a sampling bias that normalizes out across trials.

This process typically works smoothly and adaptively. Occasionally, however, runaway predictions arise as an effect of the bootstrapping structure of Bayesian sampling. This occurs when a network of hypotheses gains outsized probabilities, and contrary information is no longer able to normalize them out. Generative models of perception operate at multiple spatial and temporal scales. If the bottom-up signal at a one spatiotemporal scale does not cohere with an already-developed generative model spanning multiple spatiotemporal scales, it may be drowned out. "[E]xperiences that do not challenge a belief system become predictable and are therefore ignored" (Fletcher & Frith, 2009, p. 52).

By the time a network of hypotheses forms a coherent worldview, contrary evidence is no longer able to reweight the brain's hypotheses. "False perceptions and bizarre beliefs thus form an epistemically insulated self-confirming cycle" (Clark, 2016, pp. 80–81). The bootstrap hell is just this "self-confirming cycle," and it operates by the same principles as the bootstrap heaven that typically guides

cognition. Perceptual racial bias develops along precisely these lines.

Perceptual Racial Bias as a Bootstrap Hell

Perceptual racial biases are wide-ranging but differ from most implicit and explicit biases insofar as they directly modulate perceptual experience. Explicit biases are professions of negative associations and stereotypes, e.g., the claim that young black men are "superpredators" (DeIulio, Jr., 1996). Implicit biases are similar in content but the person may not be aware of the bias, e.g., when a judge gives harsher sentences to black defendants compared to white defendants for the same crime. Implicit and explicit biases are forms of stereotypes and are structured with semantic knowledge. What I pick out with the term "perceptual racial bias" is effects that are primarily visual and affective in nature. These effects may be linked to semantic networks but they are not themselves semantic in content. For example, subjects may misperceive objects in black men's hands as guns. While tragic instances of this in recent news are well known, laboratory demonstrations show that this is a robust phenomenon. In one experiment, participants were briefly presented with an object but were primed with a white or black face immediately before it. Priming by a black face not only increased participants' recognition of guns, but it also increased their *misrecognition* of objects as guns (Payne, 2001, 2006). Similar results were found using black participants (Kahn & Davies, 2011). In another experiment, participants in a videogame instructed to shoot armed targets were more likely to "shoot" when the target was black rather than white (Correll et al., 2002). The ecological soundness of these studies appears to be confirmed by a recent study integrating police lethal force data with regional demographic and racial bias data (Hehman et al., 2018). As with more conventional implicit biases, the presence of perceptual racial biases does not necessarily imply that the individual themselves is "prejudiced" in a conventional sense. Such perceptual biases can coexist with egalitarian worldviews.

In the terms of Susanna Siegel (2017), a philosopher of mind, such perceptual racial biases are cases of "perceptual hijacking."

> "Perception goes well, either as experience or judgment, when perceptual inputs are given proper weight. And perception goes badly, when perceptual judgment or experience is hijacked by one's prior outlook" (Siegel, 2017, p. 5).

Siegel's view of perceptual hijacking is couched in terms of epistemology and the philosophy of perception but is well suited to predictive processing. She notes that a person who is raised in a cultural milieu replete with racist beliefs about and representations of racial outgroups will naturally pick up the racism of their milieu through implicit biases—especially when that person has few meaningful interactions with the stereotyped group. The resulting racial outlook will result in hijacked perceptual experiences, such as the misperception of objects in black men's hands as guns.

Perceptual racial bias, which is a form of hijacked perceptual experience, develops in a pervasive "bootstrap hell." Initial biases in visual and other sensory information lead to predictive and generative models that are systematically biased. These initial biases are sampling biases that result from societal prejudice biases. Higher-order biases are bootstrapped up, and are reinforced by images and narratives from the prejudiced cultural milieu. These higher-order prejudice biases, in turn, affect and strengthen the perceptual biases at lower levels. The coherence at multiple spatiotemporal scales of the hierarchical generative models—an "outlook" in Siegel's language—prevents potentially contradictory and debiasing prediction errors from reweighting biased priors.

> "Cultural biases can reinforce neuronal firing patterns and result in plastic changes, reinforcing embodied practices and postures, behavioral habits, and intersubjective interaction" (Gallagher, 2017, p. 125).

In a recent study analyzing video feeds from the streets of New York, Dietrich and Sands (forthcoming) found that pedestrians move farther away from black pedestrians than from other pedestrians. This is an example of how perceptual biases and implicit biases do not exist by themselves but cohere with sensorimotor biases found even in negotiating sidewalks. A racialized outlook permeates cognition at multiple timescales, and along multiple functions, in a tangled web of hypotheses. This tangled web makes any attempt at recalibration to remove racial biases particularly fraught.

Sampling Biases in Infancy

Already at six months of age, infants express different scanning and fixation patterns for faces depending on whether those faces are same-race or other-race (K. Lee et al., 2017; Liu et al., 2015; Xiao et al., 2012). This is so early that it is presumably prior to the development of racial ingroup and outgroup *concepts*. Prior even to the development of a theory of mind, infants simulate or mirror others' mental states in the context of a "primary intersubjectivity" (Gallagher, 2017; Trevarthen, 1979).

Interaction with caretakers and with other infants sets the stage for the different facial scanning patterns based on race that we see in infants. Infants typically spend time and interact with members of their own racial group more than with members of other racial groups. Typically, this is entirely unintentional. Simply, their caretakers are more likely to be of the same racial group as the infant, as are their caretaker's friends and family. Many neighborhoods, especially in North America, are divided along racial lines. In addition, aspects of "natural pedagogy" may also be in play (Csibra & Gergely, 2009). Throughout development, we learn to perceive and attend to certain features of the world by observing and indirectly learning from our caregivers and authority figures.

This sampling bias is itself an effect of parental race as well as prejudice biases inherent in society. Caretakers may choose a daycare in an affluent area of the city where the

racial outgroup is less present, for example. Racialized housing zones and patterns of dwelling exert strong selective forces on the kinds of interactions that infants can have with racial outgroup members. For example, if the infant's daycare is in the blue zone in Figure 1, the likelihood they will be exposed to black infants is lower.

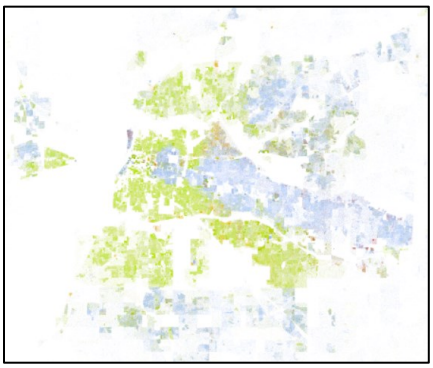

Figure 1. The racial distribution of Memphis, Tennessee. Green represents black residents; blue represents white residents. Image Copyright, 2013, Weldon Cooper Center for Public Service, Rector and Visitors of the University of Virginia (Dustin A. Cable, creator).

The embodied interactions of infants predominantly with members of their racial ingroup sets the stage for differential patterns of facial scanning and emotional processing according to race.[2] The pattern of embodied interactions predominantly with members of the racial ingroup, and the comparatively lower rate of interaction with those of the racial outgroup, creates a sampling bias for the visual system. As the visual system bootstraps its way up to hypotheses about other people's faces, the system is systematically trained on own-race faces. This changes the way own-race and other-race faces are scanned and processed. It is a kind of sampling bias that is due to the infant's social exposure: they are mostly exposed to same-race faces.

Emotion Perception and the Other Race Effect

Differential patterns of facial scanning continue to develop throughout infancy and into later life. In later development, not only do facial scanning patterns differ according to the race of the perceived face, but the emotional processing of faces also differs. Emotion perception involves vision, social cognition, and affect, and all of these can be modulated by race. Processing the emotions of racial outgroup faces is more difficult than processing those of one's own race. This leads to more attributions of negative and neutral emotionality in racial outgroup faces (Hu et al., 2017). That is, we tend to see less emotional expression, or even comparatively greater negative emotional expression, in the faces of those who are not of our own race. For example, the anterior cingulate cortex (ACC) activates for pain, but also for the empathetic perception of others' pain. In an experiment, Xu and colleagues (2009) found that when Caucasians and Asians viewed images of Caucasian and Asian faces being poked with a needle, their ACC activation increased for faces of their own race only. As we develop and gain a sense of our ingroup and of outgroups, this reinforces the differential scanning and processing patterns already in development.

The differential patterns of facial scanning and emotional processing contribute to the other-race effect. In the other-race effect, one can more clearly recognize and differentiate faces of the same race (Xiao et al., 2012). By the same token, one is less able to recognize and differentiate faces of other races. A seemingly rational inference would be that "they all look the same." Such a statement is generally perceived as harmful and prejudiced by the targets of such a statement. However, there are very real differences in how we process faces of different races, including the way we scan them, how we process their affectivity, and even our ability to differentiate them. That "they all look the same" may indeed be an accurate report of one's perceptual experience. It is an instance of "perceptual hijacking" where we cannot see the differences in other-race faces the same way that we see the differences in those of the same race.

The same bootstrapping process is involved. As small differences in processing faces develop, and as we develop different emotional reactions to these faces, higher-order beliefs such as "they all look the same" likewise develop. Just as the brain may develop a comprehensive schizophrenic or depressive worldview, integrating everything from narratives, beliefs, images, imagination, and sensorimotor habits, it equally can develop a comprehensive racialized worldview. The environment, which continues to be racially "segregated" (see Figure 1) foments interactions between racial ingroup members, while significant interactions with racial outgroup members remain statistically lower. As culturally normal racialized beliefs and outlooks become salient in later development, these in turn affect the behavior of the individual and promote a self-selection or self-segregation, even if unconscious. Explicit prejudice biases, such as narratives claiming that people of race *x* all look the same, only reinforces the lower-level visual and affective models.

Disentangling the Web of Hypotheses

In this section, I hypothesize that a consequence of perceptual racial bias is that traditional implicit bias training, or even multicultural and desegregating efforts in schools, may not be enough to eliminate them. It is not merely explicit prejudice biases, or even implicit biases, that lead to such alarming results as misperceiving objects in black men's hands as guns (Correll et al., 2002; Payne, 2001, 2006). Perceptual racial bias develops, beginning in infancy, as a "bootstrap hell" implicating multiple hierarchies of neural prediction. Facial scanning patterns and emotion perception are two examples of how perceptual racial biases develop from bottom-up data. Habituated

[2] That is not to say that this situation is the *only* reason why facial scanning patterns and emotional processing differ.

modes of perception and patterns of attention rewire neuronal connections through Hebbian learning. Moreover, the brain establishes multiple levels of neural predictions based on the priors it has developed from the person's experience in the world. The child who is overwhelmingly exposed to same-race people will naturally develop more complex ways of scanning, empathizing with, and seeing differences in same-race faces. Likewise, they will develop skewed priors (essentially, due to a sampling bias) for other-race faces.

Correcting for prejudice biases is typically done by means of implicit bias training. This is needed, especially since phenomena such as misperceiving objects in black men's hands as guns can have dire consequences (Kahn & Davies, 2011). Stereotypes, including implicit and explicit racial biases, also develop in the same process of a "bootstrap hell." "The process of picking up associations probabilistically is happening unconsciously through Bayesian principles throughout a person's life within a culture" (Hinton, 2017, p. 7). According to Hinton, stereotypes develop in the predictive brain through experience and through picking up statistical differences in the cultural world that are due to prejudice.

It may be quite difficult to disentangle the web of hypotheses, however. We may be able to shift semantic associations through promotions of multiculturalism, desegregating initiatives, and implicit bias training. Yet rewiring perceptual and affective networks may pose more recalcitrant difficulties. Through enough repetition and reinforcement, it may become challenging to perceive other-race persons differently. This can be challenging if we want to think about how to eliminate racist, sexist, and other pernicious outlooks in our society. Let us imagine a person who is forced to confront their prejudiced outlook. They begin to realize that the stigmatized racial outgroup has been the target of systematic prejudice for centuries and subsequently they relinquish the belief that members of this group are "superpredators." However, their perceptual experience will not by that fact be altered. They will continue to perceive members of the racial outgroup as threatening, even though they now metacognitively know that this is a prejudice bias. Their visual system has been plastically rewired by cultural prejudice and their perceptual experience remains hijacked.

The worry that I pose is that outlooks can induce plastic changes in the brain that could outlive the "cognitive" aspects of the outlook itself. Top-down effects recalibrate visual, attentional, and emotional processing patterns. Further complicating matters, evidence from research into infant facial scanning patterns suggests that racial biases in perceptual experience can begin to develop far before cultural beliefs, narratives, images, and other representations could possibly enter the scene (K. Lee et al., 2017; Liu et al., 2015; Xiao et al., 2012).

The interconnectedness of these elements leads to the development of a racial hyperprior. A hyperprior is a prior that is either unamenable to alteration or is only alterable under great strain. They can be innate or acquired. A simple hyperprior guiding our mundane perceptual experience is the light-from-above assumption. Because we live in a world where light typically comes from above, we interpret all images with this hyperprior (Mamassian et al., 2002). Many people likewise have been exposed to perceptual experiences of a stigmatized racial outgroup as menacing or dangerous and develop a *racial hyperprior* overdetermining their perceptual experience of members of that group. When they relinquish their narrative prejudice biases, they nevertheless persist in seeing members of the stigmatized racial outgroup as menacing or dangerous. Although the "bootstrap heaven/hell" is composed of a series of mutually-related and coherent hypotheses at multiple spatiotemporal scales and distributed across multiple regions of the brain, some of these hypotheses can go out of sync. The top-down influence of narratives of social and racial equality will contradict the bottom-up input of perceptual racial bias. Normally, higher levels of the neural hierarchy will have some effect on reweighting the priors at lower levels. The likelihood of the hypothesis (e.g., that the person in front of me is menacing) should be lower once I form an egalitarian worldview. Unfortunately, once the perceptual racial bias becomes so ingrained as to become a hyperprior it is incorrigible or at least exceedingly difficult to recalibrate (as with the light-from-above assumption).

One may no longer put full faith in their perceptual experiences once they realize that they were racially biased. However, a metacognitive doubt concerning the veracity of one's perceptual experience in many cases does not blunt that perceptual experience's emotional and behavioral impact, let alone its phenomenal appearance. In the same way, knowing that a gun is empty will not dampen the fear you feel when someone points it straight at you. Emotional effects and their behavioral consequences, at least in the immediate reaction, are not based on considered deliberation but are spur-of-the-moment processes. A person may still feel uncomfortable around a stigmatized racial outgroup even though they no longer believe that group is composed of menacing "superpredators." The racial hyperprior for perception remains when the racist outlook is removed. Years of perceptual habits have shaped their neuronal connections and neural predictions, and these reshaped connections just are the hyperprior that is so recalcitrant to reweighting. It might not be that this hyperprior is completely unamenable to intervention, but any intervention will have to be strong, sustained, and targeted. This makes racial integration a much more difficult task than we may have previously believed.

Excising culturally normal racialized beliefs and representations is insufficient to root out perceptual or attentional hijacking. Simply dropping the belief that "they all look the same" cannot by itself alter the perceptual experience of the racial outgroup's faces as self-similar or indistinguishable. More importantly, excising the belief that "they are all violent" may not in itself be sufficient to recalibrate the attentional patterns that bring about a

perceptual experience of racial outgroup members as violent.

Conclusion

A predictive processing framework has the potential to overturn our standard understanding of prejudice biases. Many racial bias interventions specifically target implicit bias. While implicit biases are undoubtably central to a racialized worldview, the latter extends across multiple cognitive, perceptual, and sensorimotor processes. Perceptual racial biases such as being primed to see a weapon, facial scanning patterns, and emotion perception in faces, and even racial biases in pedestrian movement, are among the many levels in what I have characterized as a "bootstrap hell." These levels within the bootstrap hell cohere together with narrative and imagistic prejudice biases to create a racialized outlook or worldview. Once perceptual racial biases become ingrained, however, they may become hyperpriors and hence much more difficult to eradicate. Predictive processing is still a developing field of inquiry, however, and some of the details of my account of the "bootstrap hell" of perceptual racial bias remain undertested or speculative, despite my attempt to ground it in empirical findings. If I am right, at least in broad outline if not in every detail, then we may need to rethink the way we correct for racial biases in our societies. Implicit bias training, for example, may combat the deleterious force of more cognitive aspects of racial bias. Nonetheless, perceptual racial biases as learned hyperpriors may persist in perceptual experience despite the person's best efforts to root them out.

A concrete recommendation for future studies is to disentangle perceptual and affective phenomena from the broader categories of implicit and explicit bias. Tracking the ways that cognitive implicit or explicit biases correlate with perceptual biases (i.e., visual or affective biases) may be illuminating. For example, do nonminority persons who explicitly identify as nonracist experience stress, as measured by an EEG or EMG, in the presence of racial minorities? Does implicit bias training alter stress measures in the presence of racial minorities? Does implicit bias training alter facial scanning patterns when presented with minority faces?

References

Allen, M., & Friston, K. J. (2018). From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese*, *195*(6), 2459–2482.

Avery, J. J. (2019). An uneasy dance with data: Racial bias in criminal law. *Southern California Law Review Postscript*, *93*, 28–35.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711.

Chomsky, N. (1959). A review of B. F. Skinner's Verbal Behavior. *Language*, *35*(1), 26–58.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

Clark, A. (2016). *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), 1314–1329.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*, 148–153.

DeIulio, Jr., J. J. (1996). My black crime problem, and ours. *City Journal*, *6*(2), 14–28.

Dietrich, B., & Sands, M. (forthcoming). *Using public video feeds to understand intergroup exposure on the streets of New York*.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48–58.

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11), e1000211.

Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience*, *21*, 1019–1021.

Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford University Press.

Gallagher, S., & Allen, M. (2018). Active inference, enactivism and the hermeneutics of social cognition. *Synthese*, *195*(6), 2627–2648.

Gregory, R. L. (1997). *Eye and brain: The psychology of seeing* (5th Ed.). Princeton University Press. (Original work published 1966)

Gregory, R. L. (2006). Bayes Window (2). *Perception*, *35*(2), 143–144.

Hawkins, J., & Ahmad, S. (2016). Why neurons have thousands of synapses, a theory of sequence memory in neocortex. *Frontiers in Neural Circuits*, *10*.

Hehman, E., Flake, J. K., & Calanchini, J. (2018). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological and Personality Science*, *9*(4), 393–401.

Helmholtz, H. von. (1925). *Treatise on physiological optics* (J. P. C. Southall, Trans.). Optical Society of America. (Original work published 1867)

Helmholtz, H. von. (1995). The facts in perception. In D. Cahan (Ed.), *Science and culture: Popular and philosophical essays* (pp. 342–380). The University of Chicago Press. (Original work published 1878)

Hinton, P. (2017). Implicit stereotypes and the predictive brain: Cognition and culture in "biased" person perception. *Palgrave Communications*, *3*(1), 1–9.

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

Hu, C. S., Wang, Q., Han, T., Weare, E., & Fu, G. (2017). Differential emotion attribution to neutral faces of own and other races. *Cognition and Emotion*, *31*(2), 360–368.

Kahn, K. B., & Davies, P. G. (2011). Differentially dangerous? Phenotypic racial stereotypicality increases implicit bias among ingroup and outgroup members. *Group Processes & Intergroup Relations*, *14*(4), 569–580.

Kelly, D., & Roedder, E. (2008). Racial cognition and the ethics of implicit bias. *Philosophy Compass*, *3*(3), 522–540.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge University Press.

Lee, K., Quinn, P. C., & Pascalis, O. (2017). Face Race Processing and Racial Bias in Early Development: A Perceptual-Social Linkage. *Current Directions in Psychological Science*, *26*(3), 256–262.

Lee, N. T. (2018). Detecting racial bias in algorithms and machine learning. *Journal of Information, Communication and Ethics in Society*, *16*(3), 252–260.

Liu, S., Xiao, W. S., Xiao, N. G., Quinn, P. C., Zhang, Y., Chen, H., Ge, L., Pascalis, O., & Lee, K. (2015). Development of visual preference for own-versus other-race faces in infancy. *Developmental Psychology*, *51*(4), 500.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*(4), 330–334.

Mamassian, P., Landy, M., & Maloney, L. T. (2002). Bayesian modelling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 13–36). MIT Press.

Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press. (Original work published 1982)

Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity*. MIT Press.

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), 181–192.

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

Pietraszewski, D., Cosmides, L., & Tooby, J. (2014). The content of our cooperation, not the color of our skin: An alliance detection system regulates categorization by coalition and race, but not sex. *PLOS ONE*, *9*(2), e88534.

Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research*, *41*(24), 3145–3161.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, *20*(12), 883–893.

Seth, A. K. (2015). The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. MIND Group.

Shipp, S., Adams, R. A., & Friston, K. J. (2013). Reflections on agranular architecture: Predictive coding in the motor cortex. *Trends in Neurosciences*, *36*(12), 706–716.

Siegel, S. (2017). *The rationality of perception*. Oxford University Press.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *216*, 427–459.

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. *Before Speech: The Beginning of Interpersonal Communication*, 321–348.

Xiao, W. S., Xiao, N. G., Quinn, P. C., Anzures, G., & Lee, K. (2012). Development of face scanning for own- and other-race faces in infancy. *International Journal of Behavioral Development*, *37*(2), 100–105.

Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *Journal of Neuroscience*, *29*(26), 8525–8529.