We view the world through a social lens that colors our environment with categorical labels, providing information about, among many things, people’s age, gender, and race. This lens ultimately lays the foundation for how we perceive the world and its organization, including where we live, how we make social connections, the education we receive, our healthcare, the jobs we take on, and how we go about managing our finances. Perhaps most prominently, this lens affects how we see and are seen by others. Although processing social category information may serve as an important and positive function by providing an efficient means to think about those around us, it can also have deleterious effects. Social categorization can result in the application of inaccurate stereotypes and the perpetuation of intergroup conflict.

The purpose of this chapter is to integrate across the behavioral science and neuroimaging literature on prejudice in an effort to elucidate the mechanisms of prejudice intervention from which scientists can derive innovative theoretical insights for future research. We will focus our overview and analysis primarily on racial prejudice directed toward Blacks in the United States, not because other types of prejudice do not exist, but primarily due to the unfortunate lack of available data involving other types of prejudice and groups (see the Discussion section for suggestions regarding potentially fruitful avenues for future research relating to this concern).

We will highlight the effectiveness of functional magnetic resonance imaging (fMRI) in illuminating the underlying neural substrates of prejudice. The work we review implicates a network of brain regions related to prejudice, namely those involved in person perception and emotion processing—the amygdala, fusiform face area (FFA), and medial prefrontal cortex (mPFC)—and regulation—the dorsolateral prefrontal cortex (dLPFC), anterior cingulate cortex (ACC), and orbitofrontal cortex (OFC). In addition, we present emergent evidence
for successful behavioral interventions that influence race processing across this network, with the majority of interventions explored to date altering self-regulatory processing.

Our chapter will feature answers to three questions and the theoretical implications of each for understanding the mechanisms of prejudice. We ask (and attempt to answer) the following: (1) How is prejudice defined and measured?; (2) Does a network of brain areas exist that is reliably associated with prejudice?; and (3) Are neural responses during race perception and prejudice expression malleable?

1. HOW IS PREJUDICE DEFINED AND MEASURED?

Prejudice has taken on many definitions, but for the purposes of this chapter, we regard it as any attitude or emotion toward a member or members of a group that directly or indirectly indicates some negativity or antipathy toward that group. This characterization highlights the way in which one’s social group can ultimately dictate how they are perceived and responded to. As such, one of the fundamental questions in stereotyping and prejudice research is how people extract and use information about the social groups to which targets of our perception belong. This emphasis is reflected in the field as a whole, in which a remarkable amount of research is devoted to understanding processes involved in social categorization and group identification. Of the vast array of categorical person dimensions, scholars have primarily focused on three: race, gender, and age. This is likely because each of these distinctions is readily observed from our visual appearance and each is incredibly relevant for many social judgments. Because we typically have highly accessible beliefs associated with groups that fall within these three dimensions, attending to them is also socially expedient: doing so allows perceivers to quickly and efficiently draw inferences about an individual in exchange for only minimal cognitive effort.

Although group-based associations may not be veridical or even necessarily relevant to a perceiver’s situation, they nonetheless readily come to mind. Unfortunately, these types of associations oftentimes manifest themselves as prejudices. Once social category labels are applied to someone, they can guide how a perceiver gets to know that person. For example, after a perceiver ascribes a social category label to an individual target of perception, information that is inconsistent with stereotypes about that target’s social group tends to be forgotten or explained away as unusual in nature when perceivers do not have the time, mental resources, or motivation to attend to or encode the counterstereotypic information. Pernicious associations can affect a wide range of behaviors, including subtle aspects of interactions such as nonverbal behavior (e.g., smiling or eye contact, interaction proximity), and also outright discrimination.

When most people think of prejudice, notions of harsh discriminatory acts come to mind. As such, prejudice has traditionally been characterized as explicit in nature. Explicit prejudice refers to negative attitudes based on group membership that are consciously endorsed and subject to deliberate control in their expression. This characterization connotes a high degree of intentionality in the expression of prejudice. Because of this, explicitly asking people how they feel about certain groups or members of groups, and thus acquiring their introspective reports, has been the most prevalent way to measure prejudice. Implicit prejudice, by contrast, corresponds to prejudice that (typically) lacks self-reflective access and is unintentionally triggered (c.f. Ref. 16). For these reasons, implicit prejudice is measured by performance on cognitive tasks that do not require introspection. It is thought that implicit associations are derived from affective and cognitive knowledge stored in memory acquired from years of exposure to cultural associations regarding members of social groups. These associations slowly emerge over time and unintentionally affect how we perceive and behave toward others. As a result of these well-learned associations, even individuals who are explicitly egalitarian may at times unintentionally act in prejudicial ways.

A clear distinction between trends in explicit versus implicit prejudice is possible when considering research on racial bias in the United States. On the one hand, explicit prejudice against Blacks in the US has become increasingly attenuated and is at an all-time low. Conversely, current implicit anti-Black prejudice in the US is ubiquitous. These findings highlight the fact that implicit versus explicit racial prejudice may not necessarily cohere in a way that one might expect; in fact, they may show very low to nonexistent relations.

---

*Race perception (how we come to identify the race of another person), race attitudes (the evaluations we have about a racial group), and prejudice expression (or a prejudicial behavior) are separable processes. For example, an individual can visually perceive race, but the act of race identification may not necessarily result in prejudice expression. These processes can feed off one another such that race perception can activate a host of evaluative associations based on group membership, and these associations can influence behavior. In this chapter, we will often discuss these processes together, but readers should keep in mind that the interventions discussed may not apply equally well to all of these processes. For example, interventions that affect prejudice expression may not affect race processing.*
Despite a probable lack of coherence between implicit and explicit prejudice, implicit prejudice is a rather robust phenomenon that holds implications for real-world behavior. A multitude of studies show the existence of implicit prejudice across multiple domains (e.g., gender, age, sports teams, etc.), and individual differences in implicit prejudice are predictive of discriminatory behavior. For example, several studies find that implicit prejudice predicts less friendly nonverbal behavior in intergroup interactions, as well as biased judgments in social impression formation and mock hiring decisions. In a meta-analysis of studies employing the implicit association test (IAT), a task devised to measure implicit associations between various stimuli and evaluative associations, implicit prejudice was more predictive of behaviors and judgments than explicit prejudice. The predictive utility of implicit prejudice as it relates to real-world behavior is apparent even at the societal level: Payne et al. showed that in the 2008 presidential election, voters who were higher in implicit prejudice associating Blacks with unpleasantness were either less likely to vote for Barrack Obama or rather more likely to abstain from voting altogether. The careful reader may wonder whether anti-Black or pro-Black implicit biases had a stronger impact on decisions to vote for Obama. The researchers tested this question by treating each as a separate predictor in the model. Across three studies, increasingly anti-Black attitudes predicted a lower likelihood of voting for Obama, whereas increasingly pro-Black attitudes predicted a greater likelihood of voting for him. Although these findings partially highlight the important implications pro- and anti-Black implicit biases may hold, it is also important to note that because respondents in these studies (much like many other studies on implicit race bias) exhibited an anti-Black bias on average, the net effect was a disadvantage for Barack Obama.

The findings from Payne et al. showing the relationship between implicit prejudice and presidential voting behavior serves as a strong representation of the critical relationship between implicit prejudice and population-level outcomes. Payne and colleagues’ results indicate that implicit attitudes, despite their seemingly unconscious and unintentional nature, represent a genuine and powerful roadblock to prejudice-reduction efforts at both the individual and the population levels. However, this is just one (albeit important) depiction of the relationship between prejudice and a behavior. It is important to consider a wider range of outcomes because the systematic valuation and beliefs about some groups can add to systemic oppression while the privileged position of other groups reinforces their dominance. Therefore, these and similar effects may be far-reaching and not isolated to one particular situation.

Importantly, group-based biases are not exclusive to readily apparent, deeply familiar social categories such as race, gender, and age; biased person perception is also influenced by social factors that include ostensibly incidental group membership. As an illustration of the almost inconceivable way in which this effect can unfold, the arbitrary assignment of a person to a distinct and objectively meaningless novel group is sufficient to create intergroup biases in which members of the perceiver’s own group are preferentially favored. These minimal group effects emerge implicitly, and even modulate neural responses to faces within 200 ms, implying that they occur with some degree of automaticity.

The above studies provide robust evidence that group membership is an important factor in our daily lives and affects how others within our environment are perceived and responded to. Members of one’s own social group, or their in-group, are afforded preferential attention and treatment, whereas members of the out-group are processed, on average, in a more superficial manner and treated more poorly than one’s in-group members. The magnitude of the consequences of these perceptions and behaviors can be large and can oftentimes result in prejudiced behavior, and such biases even extend to arbitrary groups with which a perceiver has little or no experience.

Many social neuroscientists, much like behavioral social psychologists, are interested in studying intergroup relations. However, they do so with a different and complementary methodological toolbox that allows for the integration of convergent evidence from an exploration of the neural systems to help understand the mechanisms that underlie these phenomena. Neuroscientists correlate neural activity within these systems with measured behaviors—such as stereotyping, prejudice, and discriminatory behaviors—to gain a better understanding of the function of these brain regions in intergroup relations. To elucidate the biological processes of intergroup dynamics, social neuroscientists have increasingly turned to functional magnetic resonance imaging (fMRI). fMRI measures brain activity by assessing the local oxygenation of neural tissues (i.e., blood-oxygenation-level dependent, or BOLD signal, which is used as a proxy measure of neuronal activity with the assumption that activated neurons increase consumption of oxygenated blood). This type of measurement is well-suited for quantifying the mechanisms of prejudice, providing insight into the psychological variables that give rise to prejudice, as well as allowing for an implicit measure of mechanism (i.e., individuals do not self-report their brain activity) that can be used to predict attitudes and behavior.

We review the relevant fMRI research on race below, first briefly considering more general functions of each brain region, and provide an overview of how each
A large body of fMRI research on prejudice implies that the expression of social group biases may share the neural circuits important for fear learning, signifying the amygdala’s potentially critical role in this domain.

A large proportion of empirical work on the amygdala’s role in social perception and evaluation demonstrates greater amygdala activity when viewing social out-group, rather than in-group, faces. For example, White perceivers show increased amygdala activity to Black faces, even in the absence of conscious awareness. The original interpretation from race perception studies was that out-group members evoke threat and consequently increase amygdala reactivity. However, this interpretation has recently been questioned as a result of inconsistencies in data, with some studies failing to report greater mean-level amygdala activity when viewing Black versus White faces for White Americans, and others finding that Black participants show either greater amygdala activity when viewing racial in-group faces or out-group faces.

Intergroup amygdala research suggests that race perception and prejudice may not show invariant effects on amygdala activation but must instead be understood within the context of the environment, perceiver, and task. Amygdala findings also support the idea that group-based amygdala differences are, in part, a function of underlying cultural associations and may be less sensitive to in-group/out-group distinctions. For example, a Black perceiver may still hold Black-danger implicit cultural associations, and so amygdala activity in this case may be greater to images of the perceiver’s in-group. Therefore, it is not necessarily the case that in the presence of an out-group target, amygdala activity will increase.

Although the interpretation that the amygdala is sensitive to cultural associations of racial groups coheres with much of the existing psychological and neuroscience literature on race perception and prejudice, the amygdala is involved in the processing of many types of salient stimuli, not just those that are negatively valenced. This points toward a need to more stringently define the parameters that affect amygdala activity in prejudice. For example, the amygdala responds to novel or ambiguous stimuli, as well as to extreme negatively and positively valenced images. Prior research also shows amygdala activity to cohere with activity in brain regions involved in motivational salience.

The amygdala may therefore function in part to inform a perceiver about what is important in the environment and then facilitate modulation of appropriate perceptual and attentional processes to respond to the salient stimulus. This implies that the interpretation of the

---

1 The FFA has largely been explored in the context of face perception. Although this region may be important for decoding and encoding person identity, there is little research that implicates the FFA in prejudice. Future research should clarify the role of the FFA in both prejudice and discrimination based on social group membership.
amygdala’s role in prejudice is likely more complex than originally thought and highlights the need for a model of the computational components of amygdala reactivity and a more nuanced determination of the predictive power of these components for discriminatory behavior. Researchers are actively moving beyond basic race perception studies to explore the relationship between amygdala activity and evaluations. Earlier, we highlighted that negative implicit associations about a social group are predictive of discriminatory behavior. Researchers examining race perception have been interested in elucidating the relationship between neural responses to racial out-group members and implicit attitudes. In an initial demonstration of this relationship, Phelps et al. had participants view pictures of Black and White faces while measuring fMRI and correlated its activity with implicit race bias. Findings showed that the greater the amygdala activity difference to Black compared with White faces, the greater an individual’s implicit (anti-Black, pro-White) race bias. However, when the amygdala is damaged, patients still display IAT (pro-White) race bias, implying that implicit associations are not strictly amygdala dependent. In addition, there is no correlation between amygdala activation and explicit race attitudes. Findings such as these support a more complex neural model of prejudice whereby these attitudes are not singularly determined, but instead involve a network of brain regions and a larger set of psychological processes.

4. RACE PERCEPTION AND EVALUATION BEYOND THE AMYGDALA

Clues about social group membership are often readily apparent when viewing only a face. As such, face perception is an important aspect of understanding how social identity is processed and recognized. A bevvy of research from social and cognitive psychology has provided strong evidence for what is known as the “cross-race effect”: individuals are faster and more accurate at recognizing faces of racial in-group, rather than out-group, members. One potential reason for this is because out-group faces are thought to be processed primarily at the category level (e.g., racial group) at the expense of encoding individuating information, and it is perhaps evolutionarily advantageous to more deeply encode in-group members. A prime neural candidate for differentiating in- versus out-group faces is the fusiform face area (FFA) in the ventral occipito-temporal cortex, as it is consistently implicated in the recognition of faces and face identity. The cross-race effect and findings showing face sensitivity in the FFA led Golby et al. to assess how this brain region relates to the in-group recognition advantage, the assumption being that the processing of in-group faces may be more nuanced and fine-grained than the processing of out-group members. This differential activity could theoretically lead to better recognition of in-group faces, mirroring the behavioral cross-race effect. In the study, White participants were asked to remember pictures of unfamiliar Black and White faces and non-face objects (i.e., antique radios). Behaviorally, the cross-race effect was replicated—participants showed superior memory for in-group White faces. Moreover, FFA activity was heightened when participants viewed same-race faces compared with other-race faces, and the in-group/out-group activation difference in the left hemisphere was correlated with the in-group memory advantage. Golby et al. reasoned that out-group members were not encoded at the individual level to an equivalent extent as in-group members, as reflected by the lesser FFA activity. This more superficial encoding may relate to poorer memory for out-group members, and future research should attempt to clarify whether the FFA is necessary to produce the in-group memory advantage.

A recent investigation using multivoxel pattern analysis (MVPA) to determine if fMRI activation patterns can predict race from face stimuli showed a much more nuanced relationship between neural activity generated in the FFA and race perception. The researchers successfully predicted the race of faces using FFA activity, but only for those who were higher in implicit pro-White bias. This finding has multiple implications. Firstly, greater bias decreases the similarity of FFA representations of race, implying that stronger race bias may be associated with larger differences in the perceptual experience of Black and White faces. That is, those who are higher in implicit race bias may show more preferential or, in this case, individuated processing, of racial in-group faces, whereas those who are lower in implicit bias may experience the perception of in- and out-group faces in a more similar manner. This finding also supports a model whereby cultural associations may drive differences in FFA activation, given that cultural associations are thought to in part drive implicit attitudes. This implies that one’s culture can shape the way even seemingly basic perceptual processing of social group membership is carried out.

Critical ways in which in-group members are distinct from out-group members are that individual perceivers typically have more experience and contact with in-group members and often assume that in-group members are more similar to them than out-group members. Research examining neural correlates of self-processing suggest that thinking about one’s own personality traits or the traits of a familiar but unrelated person (e.g., a famous actor) is linked to activity in the middle mPFC, as compared to thinking about the personality of a dissimilar person. By contrast, thinking...
about a dissimilar other results in heightened dorsal (d) mPFC activity. Recent research has applied these findings to further delineate how the mPFC distinguishes between in-group and out-group biases outside of the domain of racial prejudice. One study had participants think about the opinions and preferences of a person who had a similar or dissimilar political affiliation to their own.88 The prediction was that more politically identified participants would process the similar person as an in-group member and thus show heightened activity in brain areas, such as the mPFC, that have been linked to self-referential processing. This prediction was made for members of each separate political party, with the expectation that mPFC activity would increase in response to the perceiver’s respective political group members as compared to political out-group members. Findings showed that considering the mental state of a member of one’s own political party led to activity in the ventral (v) mPFC, whereas considering the mental state of a member of the other political party lead to heightened activity in the dmPFC. Interestingly, individuals who more strongly identified with their respective political group on an implicit measure showed heightened vmPFC activity to politically similar others and less dmPFC activity to politically dissimilar others. The interpretation of these findings is that similar others, relative to dissimilar others, are processed in a way that is closer to how the self is processed. Consistent with this hypothesis, Harris and Fiske101 showed that when participants viewed members of social out-groups that typically arouse feelings of contempt, such as drug users, less vmPFC activation occurred. If people are more able, motivated, or willing to think deeply about the thoughts and feelings of people with whom they share strong group membership, out-group members may not (on average) receive this preferential processing. However, it is also important to keep in mind that like other brain areas, the mPFC underlies a varied set of psychological processes, so one cannot always assume that self-referential processing has taken place simply due to its activation.

These findings suggest a possible differentiation in the neural correlates of in-group versus out-group perception that are relevant for the exhibition of prejudice and implicate specific subregions of the mPFC in social evaluations that distinguish between those that are categorized as similar to the self versus those that are unfamiliar or dissimilar to the self. The cited research is outside of the domain of race, and as such, the implications of this research for understanding prejudice are speculative, but some have implied that this type of in-group versus out-group processing differentiation may even form the basis of prejudice.100 Future research should seek to clarify the role of mPFC in racial prejudice.

The reviewed research thus far suggests that the amygdala, FFA (with the caveat that this may be specific to race perception), and the mPFC constitute a network that supports the representation of social group membership and evaluation. However, our perception of social category membership and our underlying evaluative associations are just one piece of the puzzle that aids in our understanding of the expression of prejudice. Although individuals readily notice social category information and are typically deeply familiar with the cultural prejudices and stereotypes that are associated with certain groups, these facts do not necessarily result in prejudicial behavior. In fact, individuals may have strong egalitarian motives that drive their behaviors in intergroup contexts. We next turn to the role of self-regulation in prejudice.

5. SELF-REGULATION AND PREJUDICE

Research on the neural regions involved in prejudice has focused on areas associated with response conflict detection and performance monitoring. This is because responding in a prejudiced manner sometimes results in a conflict between implicit associations and explicit egalitarian goals. As such, many individuals who possess chronic egalitarian goals spontaneously bring online neural mechanisms to diminish implicit race bias.102,103 This results in a conflict between biased associations and intentional response goals, resulting in activation of the ACC and dPFC. Both of these regions contribute to executive function and self-regulation, with the dPFC involved in top-down goal maintenance and emotion regulation,104,105 and regions of ACC associated with response-related selection,106 conflict detection,107,108 and inhibition of prepotent responses.109,110 This leads to the prediction that the dPFC and ACC are involved in overcoming expressions of prejudice.60,61,64 These two regions may work in concert, with the ACC detecting response conflict and the dPFC engaging regulatory mechanisms to resolve the conflict.109,111 Additionally, some perceivers may be more likely to recruit regulatory resources. For example, perceivers who hold more egalitarian beliefs may either recruit more self-regulatory resources to successfully decrease their chances of expressing prejudice or these perceivers may require less self-regulatory resources because they may have become more efficient at recruiting executive functions and/or their underlying evaluative associations may have changed. In this way, motivations to control racial prejudice may be chronically activated. A variety of studies have shown both ACC and dPFC activation in response to simply viewing out-group versus in-group faces.63,68,69,112,113 Engagement of these regions when passively viewing out-group faces may serve a preemptive function by recruiting the regulatory resources necessary for overcoming a prejudicial response.

VI. BRAIN IMAGING AND SOCIETY
Although dlPFC and ACC activation is a typical response for perceivers viewing out-group versus in-group faces, it appears that activation in these regions does not necessarily occur automatically. It is more likely that individuals must first detect the potential for racial bias in order to bring these resources online. For example, Cunningham et al. found attenuation of amygdala responses to Black faces presented when participants were aware they were viewing them (supraliminally), compared with Black faces presented when participants were unaware of their presentation (subliminal). Attenuation of amygdala activation correlated with activation of both the dlPFC and ACC. Therefore, neural mechanisms may be in place that serve to regulate racially biased responding, but only if a perceivers is aware of the potential for responding in a prejudiced manner. Studies of this nature implicate a model of prejudice regulation such that the ACC monitors for conflict between explicit intentions and implicit attitudes, and the dlPFC is brought online to control unwanted, implicit expression of racial biases.

Of course, regulatory resources such as these are not always available and are sensitive to contextual changes. For example, stress and fatigue can diminish executive functions and resource depletion exacerbates race bias, perhaps influencing the effectiveness of interventions. Moreover, situational factors can shape the egalitarian goals of perceivers and can modulate ACC and dlPFC race-based activity. For example, Krill and Platek found that being excluded by in-group partners enhanced ACC activation, as compared with social exclusion by out-group partners, implying greater conflict induced by own-race exclusion. Additionally, Forbes et al. found that listening to stereotypical music (rap compared with heavy metal) increased amygdala and dlPFC activation to Black versus White faces. These findings in concert point to the fragility of and variance in self-regulation in intergroup contexts.

It is important to note that the studies reviewed above typically involved relatively simple perceptual judgments in which race was not directly task-relevant (e.g., whether the facial stimulus was presented to the right or left of fixation). Mere exposure to a stigmatized racial out-group may activate some degree of behavioral regulation, but these studies do not provide compelling evidence regarding precisely how areas involved in behavior regulation engage in response to racial cues during more complex tasks that themselves present a regulatory challenge. Moreover, these studies do not provide direct evidence that refraining from a prejudicial response is the result of dlPFC and ACC functions. Research aimed at providing more insight into the role of prejudice regulation in more complex, real-world scenarios will prove fruitful in elucidating the mechanisms underlying its control.

The dlPFC and ACC are most likely not alone in regulating intergroup responding. The OFC inhabits the ventral (bottom) surface of the frontal part of the brain and is implicated in a variety of processes, but with respect to situations involving intergroup prejudice, it is theorized to be involved in the evaluation of the relative appropriateness of one’s responses, activating both to receiving rewards and avoiding punishments. Thus, the OFC appears to be more generally involved in current subjective evaluation. Given this association, the OFC may be a prime candidate for facilitating regulation of perceivers’ evaluations of targets when there is a potential to respond with prejudice. For example, recent research shows that OFC activity is associated with perceivers’ judgments about the potential to become friends with out-group Black individuals. Furthermore, given the OFC’s reciprocal connections with the amygdala, it may play a critical role in modulating amygdala activity elicited by exposure to racial out-group faces if a perceivers’ initial prejudiced response conflicts with explicit, overarching egalitarian motives. In this context, the OFC may function much like the dlPFC and ACC reviewed above. It will be fruitful to clarify the independent and interactive roles of the OFC, dlPFC, and ACC in integrating motives with behavior in an intergroup context. Moreover, OFC activity is associated with perceivers’ preferences for members of their own experimentally manipulated minimal groups, independent of target race. These findings suggest that the OFC may play a broader role in social evaluation, one that extends beyond the realm of racial prejudice.

To the extent that the amygdala relays information regarding expected outcomes following the perception of a target that elicits a prejudiced response, and the OFC integrates social motives with behaviors to represent the current state of the perceiver, the dense reciprocal connections between amygdala and OFC allow for a comparison of expected rewards and punishments (e.g., social shaming) with current experience (i.e., the feeling of prejudice). Support for this idea comes from research demonstrating OFC activations following value-based expectancy violations and the inability of patients with OFC damage to update representations when predictions and outcomes are incongruent. Thus, whereas subcortical systems, such as the amygdala, provide a low-resolution estimate of likely outcomes, regions of the OFC may be involved in integrating such output with current experience, allowing the current context to dictate how social evaluation is shaped.

This function of the OFC is important to consider in the domain of prejudice. It appears that the OFC functions in such situations to modulate prejudiced responses to a target so that the evaluative response coheres with the current context. Theoretically, this could result in entirely different patterns of results, depending upon
whether the experimental context is one in which social expectancies are salient versus when social expectancies are minimal, or if the environment is discouraging versus encouraging of biased responses (e.g., private versus public response conditions). Future research should vary the experimental context within which intergroup processing is measured to provide a more nuanced gauge of OFC response variability in social perception and evaluation.

The regions reviewed above are interesting to consider in the domain of prejudice, given their functional roles in emotional responding and learning, motivational salience, cognitive control, and the experience and expectation of reward and evaluation (Figure 1).

Their anatomical connections provide more reason to consider each of their roles in tandem, moving from a modulatory to a network exploration of function. Research on the neural mechanisms underlying intergroup processing provides insight into prejudice and discrimination, highlighting the basic psychological processes involved. By integrating cognitive and affective science and psychology with our social psychological knowledge of intergroup processing, we may be more likely to identify points of intervention. Translation of this knowledge into prejudice interventions requires a better understanding of not only the acquisition, storage, and expression of prejudice, but also the mechanisms that diminish prejudice.

6. ARE NEURAL RESPONSES DURING PREJUDICE EXPRESSION MALLEABLE?

Given that prejudice appears to be a rather robust and easily induced phenomenon, the fact that it can be implicit in nature, and also that a network of neural regions is reliably activated during intergroup perception and prejudice expression, should we resign ourselves to the belief that prejudice is necessary or inevitable? As behavioral and brain sciences have progressed, a more flexible view of social categorization has emerged, with dominant theory suggesting that person perception is a

FIGURE 1 The regions of the brain most commonly associated with processing of social group membership. Although these regions are involved in a number of processes, in this chapter, we highlight their theoretical contribution to racial prejudice. We propose a network of regions implicated in race processing and prejudice that includes areas important for race perception and evaluation (amygdala, FFA, and mPFC) and areas important for self-regulation (dIPFC, ACC, and OFC). The amygdala (medial view) is implicated in learning about and detecting salient things in our environments and plays a role in fear learning and expression. The fusiform face area (FFA lateral view), located in the fusiform gyrus, is thought to extract physical information from faces to distinguish among individuals. The medial prefrontal cortex (mPFC; medial view) is commonly activated when thinking about one’s self and similar others compared with dissimilar others. Together, the amygdala, FFA, and mPFC seem to support the perception and evaluation of racial out-group members. The dorsolateral prefrontal cortex (dIPFC) and anterior cingulate cortex (ACC) are implicated in self-regulation and are important for the top-down goal maintenance and emotion regulation. The orbitofrontal cortex (OFC) is involved in decision making, emotion regulation, and reward expectation and may be important for integrating societal group norms and internalized motives. The common activation of this network of regions when people think about the feelings, thoughts, and intentions of individuals from different social groups indicates that intergroup processing involves a variety of complex psychological processes.
dynamic process\textsuperscript{135} and that racial prejudice is not necessarily innate or inevitable.\textsuperscript{136,137} In fact, an effective organizational framework for prejudice intervention has been proposed. Racial prejudice is thought to contain an associative component (Automatic Prejudiced Association: Social Group X = Bad), as well as a control component (Stable Egalitarian Goal: Social Group X ≠ Bad).\textsuperscript{138} With this framework in mind, the most effective and enduring prejudice reduction techniques likely focus on altering both the automatic associative component (Social Group X = Good) as well as the initiated self-regulatory component (i.e., reinforce egalitarian goals). Social neuroscientists have recently begun to explore the neural mechanisms underlying intervention and how they relate to current cognitive and affective neuroscience models of self- and emotion regulation.\textsuperscript{44} Despite these efforts, many unanswered questions remain. We briefly review the little that is known about how the malleability of intergroup perception and prejudice is reflected in the brain. Due to the prevalence of implicit negative racial associations as opposed to explicit racial prejudice in the US, we focus our review on implicit prejudice interventions.

6.1 Counterstereotypic Imagining

Counter stereotypic imagining is a strategy that provides perceivers with concrete examples of individuals who do not conform to common stereotypes\textsuperscript{139} or prejudices.\textsuperscript{140,141} These examples can range from those who are famous or familiar (e.g., Barack Obama) or unknown and unfamiliar (e.g., a Black professor). This technique gives perceivers a counterexample that they may not otherwise encounter in their daily lives and aims to reinforce the recognition that applying overgeneralized evaluations to every individual in a social group is a flawed method. In an fMRI study that sought to understand the neural mechanisms associated with this strategy, perceivers were presented with familiar positive Black and White Americans.\textsuperscript{68} Unlike conditions in which unfamiliar individuals were presented, individual differences in amygdala activity between Black and White familiar/positive exemplars was not predictive of implicit pro-White/anti-Black bias. Recent work further delineates the amygdala’s role in counterstereotypic imagining, showing that when depictions of White and Black individuals are shown to violate stereotypic norms (e.g., a White individual in a negative role and a Black individual in a positive role) amygdala activity is heightened relative to norm-consistent behaviors.\textsuperscript{71} This finding provides additional support for the prediction that counterstereotypic imagining alters group-based amygdala processing and implies that amygdala activation may not be race- or group-specific, but instead is likely sensitive to novel and/or salient stimuli. Moreover, this work highlights the potential importance of forming counterattitudinal associations as a means of dynamically shaping intergroup processing.

6.2 Perspective Taking

Perspective taking is a prejudice-reduction technique whereby perceivers are encouraged to think about the world from the vantage point of out-group members.\textsuperscript{141} This strategy affords individuals the opportunity to understand how similar they are to out-group members and reinforces that it is important to think about the intentions and situation of out-group members, rather than relying on group-based assumptions.\textsuperscript{143} Critically, perspective taking decreases stereotyping and increases empathy.\textsuperscript{142} One possible way to attenuate differential patterns of activity observed in the neuroimaging race literature on in-group/out-group perception is to increase out-group empathy via perspective taking.\textsuperscript{101,145–150} As highlighted above, thinking of others’ internal mental states activates the mPFC,\textsuperscript{142,143} and considering the personality traits of familiar others, as compared with dissimilar others, also increases mPFC activity.\textsuperscript{97–99} The mechanism posited to account for these effects is that accessibility of the self-concept and self-other overlap in mental representations increases during perspective taking, resulting in diminished differences in mPFC activity between out-group and in-group members, and also prompting in-group-relevant processing and less stereotyping.\textsuperscript{96,141,145} These findings are relevant when considering a recent finding that shows that the neural representation of race is impacted by racial identification, or the extent to which an individual identifies with their own racial group.\textsuperscript{113} It is perhaps the case that individuals who identify less with their racial in-group view out-group members as more similar to themselves. To date, it is unclear how perspective taking shapes associative intergroup learning and self-regulatory intergroup processing. However, it is possible that having a goal to take the perspective of another person may engage greater attention to the target and self-referential processes that may relate to mPFC activity.

6.3 Individuation

Individuation is a strategy that requires a perceiver to put forth time and cognitive effort to process others.\textsuperscript{3,5} This strategy encourages perceivers to learn and consider personal information about out-group members or expectancy-violating information.\textsuperscript{171} As such, this strategy gives perceivers the ability to associate out-group members with personal information, rather than relying on group-based associations.\textsuperscript{152} Those who are similar to a perceiver, such as in-group members, are oftentimes spontaneously individuated.\textsuperscript{9} Moreover, similar to perspective taking, researchers have observed heightened...
mPFC activity when a perceiver’s goal is to form an individuated impression.\textsuperscript{144} Much research from social psychology and social cognition supports the notion that an individuation goal during an encounter can influence intergroup processing.\textsuperscript{151} In line with this, social neuroscience work has shown that when participants’ goals are to think about Black and White individuals’ preferences while viewing pictures of their faces, no race-based amygdala differences emerge.\textsuperscript{67,72} Therefore, a goal to individuate out-group targets may moderate the neural systems underlying group-based processing and perhaps increase the likelihood of more nuanced representations of out-group members.\textsuperscript{152} Future research should clarify the neural mechanism(s) underlying individuation and also consider the extent to which individuation training employing a single target generalizes to individuated processing of other out-group members.

6.4 Contact

Positive interactions with out-group members are afforded via increased intergroup contact.\textsuperscript{92,153} These types of interactions can serve to fend off negative group stereotypes and reduce any uncertainties that can arise from novel intergroup interactions. A key question in race processing is whether negative group-based associations are innate or if they rather develop due to exposure to cultural associations.\textsuperscript{154} Recent fMRI work seeking to answer this question hypothesized that amygdala differences due to race are culturally acquired and likely emerge over time.\textsuperscript{155} Results showed that race-based amygdala differences materialize during adolescence (around the age of 16) but are nonexistent in early childhood (around age 4). Moreover, greater inter-racial contact during adolescence attenuates amygdala responses to Black versus White familiar faces (Figure 2).\textsuperscript{59} However, it is important to keep in mind that simple exposure to out-group members may not be sufficient to reduce biases in amygdala activity due to race, whereas quality of contact may prove to be a more successful determinant. As an illustration of this, the number of romantic out-group partners one has is negatively associated with fear extinction learning for out-group members.\textsuperscript{154} In other words, whereas out-group members may be associated with negativity at the mean level, increased close contact attenuates this response.

6.5 Prejudice Replacement

Prejudice replacement refers to the strategy of learning to identify and replace prejudicial responses with nonprejudicial responses. For example, this could take the form of replacing the “Asians are cold” stereotype with “Asians are warm.”\textsuperscript{156} Although often described as an explicit strategy, prejudice replacement is akin in some respects to fear extinction learning. Fear extinction to in-group members is much easier than that for out-group members, suggesting that it may be difficult to restructure overly learned associations underlying out-group members, particularly in the long term.\textsuperscript{154} Recent research has extended prejudice replacement research by capitalizing on the known relationship between mechanisms involved in fear learning and expression and the learning and expression of racial attitudes, predicting that it may be possible to abolish implicit bias expression by way of pharmacological interventions that target emotional

![FIGURE 2](image-url) Panel A represents a whole-brain regression analysis exploring how childhood intergroup contact relates to amygdala activity to Black novel faces versus Black familiar faces. Cloutier, Li, and Correll\textsuperscript{59} observed greater left amygdala activity to novel Black faces compared with familiar Black faces for individuals with greater intergroup childhood contact. Panel B is a scatterplot of the relationship within the left amygdala. This research represents a remarkable advancement in our understanding of how early childhood intergroup contact can shape neural responses to race even years later. Future research should also consider the interplay between quantity and quality of intergroup contact in race processing. Panels A and B modified, with permission, from Ref. 59 © (2014) Massachusetts Institute of Technology Press.
memory and perception. An especially promising pharmacological intervention is the use of propranolol, a β adrenergic receptor antagonist that impairs memory consolidation and reconsolidation in humans. A single dose of propranolol, relative to a placebo control, effectively diminishes implicit race bias. Reconsolidation research in humans has been explored in the context of newly acquired fear associations; therefore, the mechanism by which propranolol diminishes overly learned, implicit group-based attitudes that have a complex associative structure is unknown. Although pharmacological intervention is perhaps an extreme intervention prospect for prejudice reduction, research of this type may shed further light on the psychological mechanisms underlying prejudice intervention.

The reviewed research on prejudice intervention begins to outline potential mechanisms of prejudice reduction and raises the critical question of whether there may be a common set of psychological factors that underscore these interventions that can be identified through fMRI. What mechanisms result in the most robust and reliable changes in neural activity and discriminatory behavior? Do changes within these regions predict decreases in discrimination? fMRI research in this domain has begun to address these questions. The reviewed research proposes that the most successful and lasting prejudice reduction techniques target the associative components, bolster the activated self-regulatory component, and bring processing inline with similar others (Table 1) as we observe changes both in amygdala and prefrontal activity. To date, we have not observed a change in FFA activation when manipulating these intervention techniques. However, that does not rule out a role for the FFA or other regions in implicit prejudice reduction. For example, when individuals participate in a minimal group task with mixed race participants, FFA activity is similar for in-group members of various races.

From the reviewed research, we can infer that many of these interventions influence the regions involved in forming and expressing our group-based associations, but also seem to influence areas that both regulate and integrate that information into decisions. We can also infer that this combination results in the successful attenuation of implicit biased responding, but support for this assumption is rather limited. Additionally, it remains unclear the relative importance of each of these neural regions in reducing implicit prejudice. Future work should consider under what circumstances these interventions are successful, what exact psychological mechanism(s) are altering implicit prejudice, whether these changes are lasting, and how these interventions affect real-world discriminatory behaviors. It is also unclear to what degree simply shaping race perception versus changing underlying group-based associations and/or activating motivations is impactful for reducing implicit prejudice. What is clear is that we have just begun to delve into prejudice intervention neuroscience research and future work should strive to manipulate the factors in addition to measuring individual differences that relate to these interventions. The reviewed strategies are also strikingly similar to techniques used in emotion regulation and fear attenuation and imply that researchers may benefit from borrowing theory and methods from these literature in an effort to discover more useful information about how to intervene in implicit prejudice. In doing so, however, it will be important for researchers to fully consider feasibility and external validity of these interventions as they apply to public policy recommendations. Additionally, despite the great potential for these types of programs to succeed in combating prejudice, it will also be important to fully consider the potential risks inherent in adopting them.

7. CONCLUSIONS AND CONSIDERATIONS

The understanding of social group processing and evaluation is invaluable, as it gives scholars an insight into the mechanisms involved in reducing discrimination. Social neuroscientists since the 1990s have deepened our understanding of prejudice intervention, but ultimately researchers are seeking to push this frontier further by bridging the gap between laboratory brain science and real-world behavior during interactions and judgments that hold more realistic social consequences. In other words, increased attention has justifiably been directed to the neural and psychological correlates of real-world social decision making. We do not operate in a vacuum in our everyday lives, nor should the social processes that we study in the lab. We have only begun to scratch the surface in this domain, but there is much to be optimistic about.

One promising approach to modeling how group membership influences social decision making uses neuroeconomics as a theoretical and methodological framework to bridge brain and behavioral science. Neuroeconomics is highly interdisciplinary in nature, as it combines economic paradigms, computational modeling, and neuroscience. A study by Stanley et al. (2011) was one of the first investigations to adapt a neuroeconomic approach to intergroup decision making, exploring the relationship between implicit bias and economic decisions to trust a partner. In this study, IAT was correlated with decisions to trust. Individuals with pro-White bias invested more money with White than Black partners. In a follow-up examination, greater investment in White compared with Black partners correlated with activity in the striatum, a brain region implicated in valuation. This supports
VI. BRAIN IMAGING AND SOCIETY

TABLE 1 Neuroimaging Studies Exploring Implicit Prejudice Intervention

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Strategy description</th>
<th>Study</th>
<th>Findings</th>
<th>MNI (X, Y, Z)</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Prejudice replacement</td>
<td>Identifying prejudices, labeling them, and replacing them</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>with nonprejudicial responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Counterstereotypic imaging</td>
<td>Imagining examples of out-group members who counter held</td>
<td>Schreiber and Jacoboni(^{71})</td>
<td>Greater activity to norm violating Black and White targets.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>stereotypes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Phelps et al.(^{68})</td>
<td>No longer a relationship between IAT performance and race differences in</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>amygdala activity when viewing positive famous Black and White faces</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Individuation</td>
<td>Viewing others according to their personal, rather than</td>
<td>Wheeler et al.(^{67})</td>
<td>When making an individuated preference judgment no longer observe</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>stereotypic, characteristics</td>
<td></td>
<td>race differences in amygdala activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Perspective taking</td>
<td>Adopting the perspective of an out-group member</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>5. Contact</td>
<td>Increasing exposure to out-group members</td>
<td>Cloutier, Li, and Correll(^{59})</td>
<td>Greater activity to familiar Black than familiar White faces</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Greater activity to familiar Black than unfamiliar Black faces</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Studies included in this table represent only research where these intervention techniques were manipulated. This table highlights the gaps in this research and emphasizes the need for more fMRI implicit prejudice intervention research. From this small body of research, it appears that these implicit prejudice interventions shape neural activity across a wide range of both subcortical and cortical regions.

a model whereby action values are integrated with evaluative associations in the amygdala via the striatum.\(^{164}\)

Stanley et al.\(^{38}\) study highlight’s another promising avenue for future social neuroscience research: a broader consideration of the under-examined brain structures that may be involved in social category processes. Research relevant for the striatum’s role in prejudice is scant, but its role in intergroup biases has been shown in a scenario in which perceivers have no experience with the out-group (i.e., it is experimentally manipulated).\(^{34}\) In this study, participants were randomly assigned to a novel mixed-race team
without a history of contact or conflict with an outgroup team. Subsequently, participants memorized the team membership of various faces, and these faces were then presented during fMRI scanning. Heightened ventral striatum activation to in-group, rather than out-group, faces occurred, and this activation correlated with self-reported preferences for in-group (versus out-group) members. The results from this study support social psychological posits that without a history of prejudiced responses toward the outgroup or preexisting stereotypical associations, one’s in-group may be motivationally primary. These findings expand upon the neural model of prejudice and indicate that intergroup action values may also modulate intergroup discrimination. Findings such as these that involve a wider network of brain structures will likely emerge more in the future, as the scope of fMRI and intergroup neuroscience research grows. It will be increasingly important to consider the activation of the entire network and relationships among regions to gain a richer understanding of intergroup relations.

Financial decisions represent only a subset of real-world decisions that are impacted by social category information. For example, race can also influence legal decision making, and social neuroscientists have begun to explore how neural processing of race influences judicial behaviors. Specifically, race differences in BOLD responses are shown to correlate with discrimination damage awards for Black victims, with dlPFC and parietal cortex increases relating to damages awarded. These studies provide a model for future intergroup neuroscience research to continue to bridge our laboratory research with socially consequential discriminatory behavior. Research that investigates the real-world domains where implicit prejudice is more or less likely to occur will broaden our understanding of intergroup behavior. Moreover, implicit prejudice intervention research should continue to expand into other domains in which discrimination can occur, such as education, employment, and health care.

There is a clear need for more work outlining the psychological and neural factors of implicit prejudice intervention. Initial studies have focused on basic-level phenomena or behavioral processes that exacerbate implicit prejudice. More recent work is extending beyond exploring only the factors that produce and exacerbate racial bias to understand the mechanisms of implicit prejudice mitigation. With these efforts, a more detailed picture of the underlying psychological and neural mechanisms of prejudice intervention will emerge. Additionally, it is not enough to demonstrate diminished prejudice in the lab, and research should aim to extend our understanding of how the neural and behavioral correlates of prejudice intervention predict real-world decreases in discrimination. A possible next step in this area is to compare the impact of intervention strategies and explore how effective these strategies are across contexts to provide a better understanding of which strategy is most effective, and in what types of situations.

To date, the majority of the prejudice literature examines responses to Black and White race categories in US participants, and as a result, this chapter primarily concentrated on this literature. It is important that future neuroscience work on race include a variety of racial groups across cultures to facilitate a more complete understanding of stereotyping and prejudice and the steps that can be taken to diminish discrimination. Moreover, researchers interested in social group membership should broadly sample both in terms of stimuli and participants when exploring discrimination interventions to make more informed policy recommendations. By combining affective, social, cognitive, and economic neuroscience approaches and insights with decision tasks reflecting socially relevant consequences, we will obtain a better understanding of how our implicit biases may, or may not, impact the choices we make. Additionally, we highlighted some studies that employed the use of novel groups (e.g., Ref. 34). Studies such as these allow for the examination of biases in social processes that may generalize across social groups, those that may not be confounded with preexisting attitudes and cultural associations, and provide insight into the basic set of psychological and neural processes that underlie discrimination. We view these types of studies as critical to understanding the perception and processing of social categories more generally, and as such, an important foundation from which more nuanced examinations may be conducted.

References


33. Krumbarb K, Corinelle O. Holistic processing is tuned for in-
VI. BRAIN IMAGING AND SOCIETY


