Us and Them: Cognitive and Neural Mechanisms of Intergroup Behavior

Maddalena Marini and Mahzarin R Banaji, Istituto Italiano di Tecnologia (IIT), Center for Translational Neurophysiology of Speech and Communication, Ferrara, Italy; and Harvard University, Department of Psychology, Cambridge, MA, United States

© 2022 Elsevier Ltd. All rights reserved.

Introduction 509
Group Membership: A Social Categorization Process 510
Effects of Group Membership 511
Implicit Social Cognition 511
Perceiving and Understanding Others 513
Conclusion 514
References 514

Glossary

Default Mode Network (DMN) System of regions identified during rest, i.e., when the brain is not engaged in any specific mental task. Core areas of DMN include the medial posterior cortex – specifically the posterior cingulate cortex (PCC) and parts of the precuneus, medial prefrontal cortex (MPFC), bilateral inferior parietal lobule (IPL), and posterior temporal areas around the temporoparietal junction (TPJ). Interestingly, several regions of this network are also active during self-referring processing and metalizing, i.e., cognitive processes of inferring people’s thoughts and feelings. This latter finding led social neuroscientists to suggest that DMN plays a role in social cognition and intergroup behavior.

Discrimination Negative behavior toward individuals because of their group membership; also, negative behavior toward groups because of their groupness.

Empathy The ability to understand and experience another’s feelings and emotions.

Galvanic Skin Response (GSR, also named skin conductance response, SCR) Refers to the changes of sweat gland activity in the skin, which is controlled by the sympathetic nervous system. It reflects a measure of emotional arousal, i.e., the intensity of an emotional state.

Implicit Association Test (IAT) Behavioral instrument to measure the automatic strength of association between concepts (e.g., insect/flower or dark skin/light skin) and attributes. When the attributes consist of global evaluative attributes such as good/bad, pleasant/unpleasant, positive/negative, the resulting association is called an implicit attitude. When the attributes consist of specific traits, e.g., tall/short, smart/dumb, honest/dishonest, the resulting association is called an implicit belief or implicit stereotype. When the concept represents self/other, the resulting association refers to as implicit esteem (if the attribute dimension is good/bad) and implicit identity (if the attributes represent specific qualities such as traits).

In-Group/Out-Group Membership and Identity In-group is the term used to refer to groups of which one is a member. Out-group is the term used to refer to groups of which one is not a member. Group Membership is the term used in an all-or-none sense to refer to whether one is a member of a group or not, e.g., one is Brazilian or not, a 7th grader or not. Group Identity is the term used to capture the variation in the strength of the individual-group relationship. Two individuals may both be members of the group “Catholic”, but one may be more strongly identified as such than the other. Group identity is the term used to refer to this varying strength of individual-group attachment.

Intergroup Behaviors Actions between individuals or groups whose origin can be traced to group membership.

Methods Used in Social Neuroscience Social neuroscience employs several instruments and methods to investigate the associations between biological and social processes, including functional magnetic resonance imaging (fMRI), event-related potentials (ERPs), noninvasive brain stimulation (NBS), and galvanic skin response (GSR; also known as skin conductance response, SCR). Functional magnetic resonance imaging (fMRI) measures brain activity by detecting regional and time-varying changes in brain metabolism and blood oxygenation. This technique relies on the fact that the change of oxygenated versus deoxygenated blood flow increases when a brain area is active. These alterations are detected by using a magnetic field. It can be used to produce activation maps showing which parts of the brain are involved in a particular mental process. It has an excellent spatial resolution. Event-related potentials (ERPs) measure electrical activity generated by the brain in response to specific internal or external events (e.g., sensory, cognitive, or motor stimuli). Electrical activity is detected using electrodes placed on the scalp and amplified through an electroencephalography (EEG) machine. It is employed to examine brain functioning, psychophysiological correlates, and timing of mental processes. It has an excellent temporal resolution.

Noninvasive brain stimulation (NBS) is a method that alters the ongoing neural activity in given brain areas and distributed networks. Unlike the traditional neuroscientific techniques (e.g., ERPs and fMRI), NBS allows researchers to investigate brain areas that are causally related to a given behavior or to modulate their activity. One of the most popular NBS tools is transcranial magnetic stimulation (TMS). Transcranial magnetic stimulation (TMS) induces small electric currents in the brain to change...
Human individuals are, from the moment of birth and through their life course, members of social groups. They identify as members of particular groups with which they share some similar characteristics, they act to protect and favor their associates, and they behave and evaluate others based on their group membership and identity (Tajfel and Turner, 1979, Ashburn-Nardo et al., 2001; Otten and Ventura, 1999; Van Ravel and Cunningham, 2009, 2011).

Individuals’ knowledge of belonging to certain groups can be defined as social identification (Tajfel and Turner, 1979) and reflects the process that leads people from an individual (i.e., “I”) to a collective recognition of themselves as a group (i.e., “we”) and to the creation of a different representation of themselves than members of other groups (i.e., “us” and “them”) (Brewer and Gardner, 1996). This distinction between members of one’s own group (i.e., in-group members) and members of the other groups (i.e., out-group members) is based on an initial process of social categorization. We categorize people into categories based on some similarities or something in common to instantly conceive the world and people around us and define our interaction and behavior with others (Bruner, 1957).

Group membership has important effects on individuals’ behavior. That is, people in groups tend to display favoritism toward in-group members and negative stereotypes, evaluations, and behaviors against out-group members. This propensity to favor in-group members has been observed in every culture on earth (Brown, 1991) and in children as young as five (Bunham, 2011).

In the last decades, increases in globalization have brought issues of multiethnic and multicultural diversity in many societies to the fore and raised scientific interest in questions concerning the cognitive and biological processes that underlie intergroup relations (Eberhardt, 2005; Ito and Bartholow, 2009; Kubota et al., 2012; Marini et al., 2018) among others. The goal has been to understand the mechanisms of group representation and group-based preferences in behavior and in the brain and to understand their degree of malleability. Can life-long learning that a particular group is “bad” (and by extension has qualities like untrustworthy or incompetent) be transformed into a more neutral valence? An interdisciplinary approach—captured by the term social neuroscience (Cacioppo et al., 2000) – has led to an understanding of mind and behavior via an integrative analysis that combines social and biological theories and methods (Cacioppo et al., 2010).

In this article, we provide an overview of studies on intergroup cognition observed in behavior and brain mechanisms. In particular, we focus on research investigating the mechanisms underlying intergroup behaviors and the effects of group membership and social categorization on cognitive and neural processes that support social interactions with others. In doing so, we hope to bring
attention to the importance of these findings to understand intergroup behaviors by investigating cognitive processes in social context.

**Group Membership: A Social Categorization Process**

The process of social categorization can be considered the first step in understanding the psychology of intergroup behavior, as all future thoughts, feelings, and actions depend first on the ability to place individuals accurately into perceived classification by group and only if our minds are capable of representing groups as separate entities. According to Tajfel and Turner, social groups can be defined as "a collection of individuals who perceive themselves to be members of the same social category, share some emotional involvement in this common definition of themselves, and achieve some degree of social consensus about the evaluation of their group and their membership of it" (1979, p. 40).

Social groups can be formed based on existing social categories that are highly salient and pervasive in our society - such as race, age, and gender - but they can also be established virtually via any perceivable distinction between people or specific factors in the social context, including the presence of competitive or cooperative functional relations between individuals (Cikara and Van Bavel, 2014). In other words, groups are so fundamental to human perception and social cognition that even a minimal distinction based on a coin flip is sufficient to create in-group favoritism and/or out-group derogation. They can even override robust group membership associated with pre-existing social categories (e.g., race) or previously established functional relations (Tajfel, 1970; Tajfel et al., 1971; Locksley et al., 1980; Van Bavel and Cunningham, 2011; Gutsell and Inzlicht, 2013). This phenomenon has become known as the "minimal group paradigm".

Studies that directly interrogate the brain have shown that social categorization is a spontaneous process (Crisp and Hewstone, 2007) that rapidly occurs in the brain, within about a quarter of a second (Ito, 2011). For example, Ito et al. (2004) measured event-related potentials (ERPs) in response to faces of different races and showed that at around 250 ms human brain distinguished between in-group and out-group members. That is, larger electrophysiological brain activity was observed in response to faces of racial in-group than out-group members. These early neural differentiation effects of in-group and out-group were confirmed in both White and Black participants (Dickter and Bartholow, 2007) and also revealed in other studies presenting male and female faces. In this latter case, distinct responses for faces of different gender occurred within 200 ms (Ito, 2011), suggesting that some social categories may be processed and distinguished before others. Similarly, different early brain responses to in- and out-group members have been observed even when social groups are experimentally formed using a minimal group paradigm. For example, Ratner and Amodio (2013) examined the electrophysiological brain response to faces of in-group and out-group members in a study in which participants were arbitrarily assigned to two groups based on a bogus dot estimation task. In this study, participants were first asked to estimate the number of dots in various patterns and then randomly labeled as "underestimators" and "overestimators". Results showed greater N170 ERP peaks when participants were presented with faces of people that belong to their own group, indicating that the brain differentiates in-group and out-group members even when social categories are not highly salient or marked by distinct visual cues (e.g., race and gender). They can emerge even if defined in a simulated and arbitrary way, suggesting the neural, cognitive, and social sovereignty of social groups as guides to thought and behavior.

Similar brain responses using pre-existing or new experimental groups have also been found in a series of functional magnetic resonance imaging (fMRI) studies, indicating that social categorization is associated with the activity of specific brain areas involved in affective and emotional processes. Consistently, studies have reported greater amygdala activity when participants are presented with other-race faces than own-race faces (Hart et al., 2000; Phelps et al., 2000; Richeson et al., 2003; Cunningham et al., 2004a; Wheeler and Fiske, 2005; Ronquillo et al., 2007; Kubota et al., 2012; Stanley et al., 2012). Also, when they view faces of in-group than out-group members defined using minimal-group paradigms (Van Bavel et al., 2008). The amygdala is a small subcortical region in the anterior-temporal lobe implicated in fear conditioning processes (e.g., fear learning and expression) and detection of emotionally relevant stimuli in the environment (e.g., negative stimuli) (Davis and Whalen, 2001; Hamann et al., 2002; Anderson et al., 2003; Phelps, 2006). These findings have led scientists to suggest that social categorization appears prepared to automatically associate negative affect to members of out-groups. In particular, fMRI studies using minimal group paradigms have provided evidence that different neural responses to in-group and out-group members are not due to prior experience with the group or to its salience. For example, Van Bavel et al. (2008) randomly assigned subjects to either “Leopards” or “Tigers” mixed-race teams, i.e., teams composed both by White and Black members and instructed them to identify faces of individuals, both Black and White, who belonged to their own team versus the other team. Results showed an increase in amygdala activity, along with other brain regions (i.e., orbitofrontal cortex and fusiform gyrus), when participants viewed faces of in-group than out-group mixed-race members. In other words, Black individuals who were members of one’s own team were quickly viewed as in-group with neural patterns indicating such inclusion. These results indicate the flexibility of the neural system to learn new groupings of “my team” and in opposition to older learning of White and Black as traditional in- and out-groups (see also, Ratner and Amodio, 2013; Van Bavel and Cunningham, 2009).

Taken together, these studies show that social categorization of individuals into groups and social group distinctions are automatic processes marked by spontaneity, rapid brain responses, and engaging affective and emotional responses. Most importantly, and even optimistically for the future of intergroup relations, this phenomenon is not exclusively determined by pre-existing and salient social categories that have been traditionally present in society but can be shaped to include a diversity of individuals as a part of one’s own or other groups.
Effects of Group Membership

Group membership plays a key role in our interactions with others beyond social categorization, to affect downstream cognitive and neural processes that support unique patterns of intergroup behaviors.

Group membership generates in-group favoritism, stereotyping, negative attitudes or affect toward others, as well as explicit forms of prejudice and discrimination (Tajfel, 1970; Tajfel et al., 1971). These phenomena are inevitable products of social categorization (Allport, 1954). Categorizing people into social groups, in addition to represent an immensely useful human ability to conceive the world in terms of categories, can also lead us to draw inferences about others based on our knowledge of the categories to which they belong and drive the display of negative evaluations or biased behavior toward individuals belonging to specific social groups.

In addition, group membership influences low and high-level cognitive processes underlying intergroup behaviors, including how we perceive, recognize and represent others, as well as our understanding of their thoughts and feelings (Adams et al., 2010; Golby et al., 2001; Ito and Urdland, 2003; Kelley et al., 2002; Mitchell, 2009). These cognitive effects are reflected both in neural and electrophysiological activity, as well as in behavior.

Implicit Social Cognition

For decades, social and cognitive scientists have studied stereotyping and attitudes (called "prejudice" in the intergroup context) to understand the mechanisms that sustain it. These studies showed that stereotyping or evaluating individuals based on the fallible assumption that certain characteristics (e.g., goodness, trust, intelligence) are universally true of all group members can be inaccurate and self-perpetuating (Snyder and Swann, 1978; Harris and Rosenthal, 1985; Stangor and McMillan, 1992; Steele and Aronson, 1995; Stone et al., 1997; Fiske, 1998), leading to misjudgments (Tajfel and Wilkes, 1963; Hamilton and Gifford, 1976; McGarty and Turner, 1992) that can resist disconfirming evidence (Kunda and Oleson, 1997; Reuben et al., 2014).

Importantly, research on stereotypes and attitudes revealed that these processes can operate with relatively little conscious awareness or conscious control over the behavior. This discovery had profound influence as it raises the question of discriminatory behaviors occurring without the actor’s knowledge, and often in opposition to the actor’s own conscious values of fairness and egalitarianism (Banaji and Greenwald, 2013).

Research on stereotypes and attitudes that occur outside of conscious awareness has been captured by the term implicit social cognition (Greenwald and Banaji, 1995). Specifically, implicit social cognition is the study of cognitive processes and individuals’ mental representations concerning three psychological constructs: implicit attitudes, implicit stereotypes, and implicit identity (Greenwald and Banaji, 1995).

Implicit social cognition relies on indirect measures to capture psychological constructs that are not accessible to consciousness. Unlike explicit measures (e.g., self-reports) in which the content is assessed directly, implicit measures infer constructs of interest through behavioral performance (e.g., individuals’ response latency or errors in specific tasks). One of the most widely used measures of implicit social cognition is the Implicit Association Test (IAT) (Fig. 1; Greenwald et al., 1998).

Although implicit and explicit measures tend to show positive correlations, they can display a large degree of dissociation (e.g., age attitudes; Nosek and Smyth, 2007; Nosek et al., 2007; Greenwald and Nosek, 2008) or even opposite effects (e.g., weight attitudes; Marini et al., 2013; Marini, 2017). These findings suggest that implicit and explicit measures assess related but distinct constructs (Nosek, 2007). For example, a study conducted in a large sample showed that while White individuals endorsed egalitarian views at the explicit level and pro-White attitudes at the implicit level, implicit and explicit measures were moderately positively correlated, indicating some shared and independent variance (Nosek et al., 2007). Moreover, Cunningham et al. (2004b) used structural equation modeling to show that although implicit and explicit intergroup attitudes are quite highly correlated (~0.50), they cannot be modeled as a single factor and thus considered as unique constructs. This result cannot easily be explained as a function of method variance (i.e., that the methods to tap implicit and explicit intergroup attitudes vary greatly), because implicit attitudes unrelated to implicit intergroup cognition (e.g., implicit insect-flower attitude) do not cohere with implicit intergroup attitudes.

Implicit social cognition research has demonstrated that implicit attitudes and stereotypes are pervasive (Nosek et al., 2007; Sabin et al., 2012; Marini and Banaji, 2020) and predict variations in behavior across a variety of topics, in many cases above and beyond explicit measures (Fazio et al., 1995; Dovidio et al., 1997; Greenwald et al., 2009) (a meta-analytical comparison of the predictive power of implicit and explicit measures is given in Kurdi et al. (2019)). For example, it has been shown that implicit pro-White attitudes predicted physicians’ decisions to provide more thrombolysis recommendations for White than Black patients with acute coronary syndromes (Green et al., 2007), nations with stronger implicit stereotypes associating science with male than female predicted a greater national-level gap between women and men in science and math achievement (Nosek et al., 2009), and implicit identity associating self with death predicted suicidal ideation and the probability of making a suicide attempt in the next six months (Nock et al., 2010). Recently, a concept captured by the term “the bias of crowds” has been used to show that implicit attitudes, obtained at the aggregate level, reliably and significantly predict behavior (Payne et al., 2017): e.g., aggregate IAT data today, represented at the macro level by states within the US, reflect earlier proportion of slaves to non-slaves. That is, states with higher IAT race bias today are also the states that had larger slave populations in the past (Payne et al., 2019). As another example, aggregate IAT race bias by area predicted lethal use of force by police (Hehman et al., 2018).
In addition, studies examining the malleability of implicit social cognition have shown that implicit attitudes and stereotypes are relatively stable. That is, although they can be immediately shifted in response to brief interventions, these effects do not last over time (Dasgupta and Greenwald, 2001; Marini et al., 2012; Lai et al., 2016). For example, a study with a large sample size showed that even the most effective interventions to reduce implicit race attitudes (Lai et al., 2014) produced short-term changes that persisted only after a delay of one or a few days (Lai et al., 2016). However, we now have the first demonstration that implicit attitudes can change over the long-term. Analyzing aggregate data over a 10-year period and using time-series analyses, Charlesworth and Banaji (2019) showed that implicit attitudes can change in a stable and predictable manner with attitudes toward sexuality (gay-straight) being the striking example. Why such change has occurred is not yet known, and future research is needed to understand whether societal changes (e.g., legislation, media representation of specific social groups) might be the drivers of attitude change at an individual level.

With the application of neuroscientific techniques, researchers showed that implicit social cognition involves brain areas associated with the processing of emotion and executive functions (Phelps et al., 2000; Richeson and Ambady, 2003; Cunningham et al., 2004a). For example, Phelps et al. (2000) found that individual differences in implicit racial attitudes were correlated with the activity of the amygdala. That is, White participants with stronger implicit pro-White attitudes showed greater activity in this subcortical area when presented with Black than White faces. Because the crucial role of the amygdala in the acquisition, storage, and expression of fear conditioning and unpleasant emotions (e.g., disgust) (Phelps and LeDoux, 2005), scientists have suggested that these results reflect negative evaluations toward other-race members based on individuals’ cultural and historical social learning, past personal experience and own group membership (Phelps et al., 2000; see also Senholzi et al., 2015).

Similarly, Richeson et al. (2003) found that stronger pro-White attitudes were also associated with greater activity of the anterior cingulate cortex (ACC) and dorsolateral prefrontal cortex (DLPFC). The ACC is a brain area involved in detecting and monitoring cognitive conflict, i.e., when automatic and intentional responses are in competition, while the DLPFC plays a critical role in implementing cognitive control and regulation processes (Carter et al., 1998; MacDonald et al., 2000; Barch et al., 2001; Botvinick et al., 2001).

These findings have led researchers to suggest a role for the experience of conflict between implicit and conscious cognition, such as intentions to respond fairly. That is, people may seek to control their implicit attitudes or stereotypes because they are socially unacceptable or are in conflict with their conscious beliefs of fairness and egalitarianism. To test this hypothesis, Amodio et al. (2004) measured the magnitude of the error-related negativity (ERN), an electrophysiological component that is elicited when errors occur, in people who reported a desire and motivation to be free of prejudice. In this study, participants were asked to complete a priming task assessing their implicit stereotype associating Black people with guns (i.e., violence). That is, they were asked to rapidly categorize a series of images representing guns or non-violent tools, by pressing two response keys; each image was primed with the face of a Black or White person. Results showed that participants were more likely to inaccurately classify the image of a tool as a gun when this behavior was primed with a Black than a White face, indicating thus the presence of an implicit stereotype linking Black Americans with guns. In addition, these mistakes were associated with a higher ERN component compared to those related to the White prime (i.e., when the image of a tool, primed with the face of a White person, was wrongly classified as a gun).
classified as a gun) and a slowing down of the performance in the next trials of the task. These results provide evidence that implicit stereotypes are lawfully associated with the brain’s electrophysiological activity evoked during error processes and support the hypothesis that people try to overcome their automatic racial associations by avoiding additional biased errors.

In sum, these studies show that we make inferences about people based on the social categories or groups to which they belong. This process is implicit and automatic (i.e., it can occur without awareness and intentional control), and it can lead us to misjudge others, generating prejudice and discrimination. Implicit attitudes and stereotypes involved brain areas and electrophysiological mechanisms associated with the processing of emotion and executive functions.

**Perceiving and Understanding Others**

Extensive research has shown that group membership modulates activity and responses of specific brain regions and electrophysiological components that play a crucial role in lower-level cognitive processes underlying intergroup relations. For example, group membership can generate a perceptual bias in favor of in-group than out-group faces. That is, people recognize the faces of members of their own racial group better than faces of other racial groups (Feingold, 1994; Meissner and Brigham, 2001). This perception bias, known as own-race or other-race effect, is associated with the activity of the fusiform face area (FFA), a brain region in the ventral visual processing stream that is specialized for facial perception and recognition (Kanwisher et al., 1997; Rhodes et al., 2004) and is shown to be more active when people are presented with own-race faces than other-race faces (Golby et al., 2001).

In other words, a face feels more like a face when it represents a member of one’s own group and may call out subsequent cognitive activity to humanize such an entity more than out-group members. These results have also been replicated using minimal group paradigms, indicating thus that the enhanced recognition of in-group than out-group faces is not due to prior experience with the own-race group (Bernstein et al., 2007; Van Bavel et al., 2011; Van Bavel and Cunningham, 2012).

Similarly, it has been shown that the presence of an in-group perceptual bias also influences the recollection of information, events, and facts associated with specific social groups (social memory; Hugenberg et al., 2010, for a review). For example, Molenberghs et al. (2013) conducted a fMRI study in which participants were asked to judge the speed of actions performed by in-group or out-group members. Results showed that participants judged the actions of in-group players as faster than actions performed by out-group players, although on average both groups had similar performances. This bias in perception was associated with activity in the left inferior parietal lobule, a region that is implicated in transforming the representation of actions from a visual to a motor program (Rizzolatti and Craighero, 2004; Rizzolatti and Fabbri-Destro, 2008). Readers should appreciate that even behavior like speed of movement in a neutral context (i.e., not warfare) that one might imagine would yield to objective analysis is shaped by knowledge of group membership.

In addition, studies also show that group membership influences not only lower-level cognitive processes but also plays a critical role in shaping higher-level neural operations, including how we understand the thoughts and feelings of others. For example, studies show that humans are more accurate when inferring the mental states of own-race than other-race group members (Adams et al., 2010) and that this effect may arise from perceived similarity of in-group members to self. Indeed, we know that greater activation of the medial prefrontal cortex (mPFC), a core brain region of the default mode network (DMN) involved in self-referential processing and mentalizing (Kelley et al., 2002; Mitchell, 2009), is observed when a task requires representation of in-group than out-group members (e.g., Mitchell et al., 2006; Volz et al., 2009; Jenkins and Mitchell, 2011). For example, in a recent fMRI study, Morrison et al. (2012) showed that in-group labels defining broad social concepts (e.g., gender, nationality, or religious affiliation) elicited increased activation of the mPFC compared to out-group labels. Similar results have also been found using a minimal group paradigm, in which participants were asked to make evaluative group decisions by dividing money between in-group and out-group members (Volz et al., 2009). That is, greater activation of the dorsal mPFC was observed when subjects had to choose between allocating money to an in-group than an out-group member, compared to decisions involving two in-group or two out-group members.

Similarly, neuroscience research provides evidence that social membership influences specific patterns of brain activity associated with an emotional state described as empathy, i.e., the ability to understand how another person feels, including pain and distress (for reviews, see Batson, 2009; Walter, 2012; Zaki et al., 2012). In particular, these studies showed a decrease in neural and physiological empathic responses when people view members of other social groups in physical pain compared to viewing members belonging to their own social group. For example, Xu et al. (2009) conducted a fMRI study in China where White and Chinese participants were presented with pictures of both White and Chinese faces receiving either painful (i.e., face pricked by a needle) or harmless stimulation (i.e., face touched by a Q-tip). Results showed greater activation of the anterior cingulate in response to painful versus nonpainful stimulation only when the face in the picture matched the participants’ own racial group. The anterior cingulate cortex (ACC), in addition to being involved in cognitive conflict processes, has also been implicated in the so-called “pain matrix”. The ACC, together with the supplementary motor area and the insula, are considered to represent the cerebral circuit for pain perception (Decety, 2011; Eres and Molenberghs, 2013). Similarly, Avenanti et al. (2010) used transcranial magnetic stimulation (TMS) over the motor cortex to measure the corticospinal excitability to other’s pain. They showed that the empathic sensorimotor reactivity (i.e., the reduction of the motor corticospinal excitability observed when an individual watches a video of another hand being pricked by a pin, Avenanti et al., 2005, 2006) was present only when participants watched a video of an in-group hand (or even an artificially colored, purple hand) being pricked by a pin but not when the hand belongs to an out-group member. The lack of empathic sensorimotor reactivity was associated with stronger implicit racial bias measured by means of the IAT. Consistent with these results, Forgiarini et al. (2011) showed that even skin conductance response (SCR)
Us and Them: Cognitive and Neural Mechanisms of Intergroup Behavior

(i.e., an index of automatic arousal of the nervous system) is lower when White participants are presented with images of Black targets’ painful experience compared to when they view White targets in the same painful situation. Interestingly, different brain responses to in-group and out-group members’ pain have also been found in competitive contexts, such as among sports fans. In an fMRI study by Cikara et al. (2011), fans of the rival Yankees and Red Sox baseball teams watched videos in which either their team or the rival team succeeded. Results showed increased activity in the insula, a region included in the “pain matrix” (Decety, 2011; Eres and Molenberghs, 2013), when fans watch their team fail to score, and when they watched the rival team perform well.

Taken together, these findings show that intergroup situations create perceptual, cognitive, and affective responses that indicate the importance of group existence. That is, investigations of both behavior and brain have shown systematic differences in processes that are engaged when encountering members of our “tribe” or not.

Conclusion

In the present article, we discussed research investigating the cognitive and affective processes underlying intergroup categorization, perception, and evaluation that underpin the behavioral and brain mechanisms that support the psychological dynamics of intergroup behavior.

Research reviewed provides evidence that intergroup behavior is based on an initial process of social categorization that occurs spontaneously and rapidly, as measures of brain activity and behavior demonstrate. The network of brain activation includes the social cognition network that overlaps with the default network and regions known to be engaged in the processing of affective and emotional responses. Categorizing people into groups influences all crucial cognitive processes that define social evaluations and interactions. It determines how we perceive, recognize, and represent others, as well as how we infer their mental and emotional states. In particular, we know that these processes are biased in favor of individuals perceived to be members of our own social groups, whether they are groups like race and nationality, sports teams, or entirely arbitrarily created for experimental purposes. Cognitive processes such as better memory for in-group members and more accurate inference of the thoughts and feelings of similar others are well-established. Moreover, social categorization underpins implicit attitudes and stereotypes that operate with relatively less conscious awareness or conscious control to produce patterns of discrimination that can go unnoticed and unchecked.

However, the studies reviewed also point to social categorization as a flexible and highly context-dependent process. The sense of belonging to a specific social category can indeed be shifted by the demands of the situation or environment, such as requiring new ways of defining the in-group, or by inducing cooperation or competition between groups. Malleability is visible in surprising ways when newly specified categorizations are shown to override pre-existing salient social categories, pointing out the adaptive nature of the social cognitive foundation of intergroup processes.

These findings have relevant implications for reducing intergroup conflict. This research to date underscores the idea that by changing the social context and its dynamics, it is possible to modulate the relevance of different social categories (Turner et al., 1987) and thus also their effects on behavior. Over the past few decades, psychologists have proposed several behavioral interventions to reduce intergroup conflict, e.g., by shifting group boundaries or group affiliations through competition or cooperation, inducing emotional states or egalitarian goals, and exposing individuals to counterstereotypical exemplars or information (see for example, Dasgupta and Greenwald, 2001; Blair, 2002; Gawronski and Bodenhausen, 2006; Olson and Fazio, 2006; Marini et al., 2011, 2012; Lai et al., 2014, 2016).

In addition, these studies have also allowed researchers to develop interventions that produce interference in the activity of specific brain regions associated with intergroup conflict and stereotyping. For example, recent studies using noninvasive brain stimulation (NBS), a method that interferes with ongoing brain activity in targeted cortical areas and distributed brain networks (Rotenberg et al., 2014), showed that by modulating the activity of the areas that are causally involved in intergroup behavior (e.g., anterior temporal lobe, ATL; medial prefrontal cortex, mPFC), it is possible to shift stereotypes and negative attitudes toward specific social groups (for a review; Marini et al., 2018).

Finally, the studies we reviewed provide evidence that a multidisciplinary approach that combines behavioral and neuroscientific methods and theories, can provide a more comprehensive understanding of intergroup behavior and demonstrate reductionist coherence, i.e., that the analysis that emerges at one level (behavioral) is substantiated by analysis at another level (brain activity). Intergroup behavior is a product of many forces, from evolutionary pressures that made social groups ubiquitous in the evolution of humans and other animals, to the long histories of specific intergroup relations, to an individual’s own biological and social development, as well as the demands of the immediate situation.

References


