Journal of Experimental Social Psychology 83 (2019) 55–69

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Sudden shifts in social identity swiftly shape implicit evaluation

Yi Jenny Xiao[a,*], Jay J. Van Bavel[b,c]

[a] School of Interdisciplinary Arts and Sciences, University of Washington Tacoma, 1900 Commerce St, Tacoma, WA 98402, United States of America
[b] Department of Psychology, New York University, 6 Washington Pl, New York, NY 10003, United States of America
[c] Center for Neural Science, New York University, 4 Washington Pl, New York, NY 10003, United States of America

ARTICLE INFO

Handling editor: Kimberly Rios

Keywords:
Social identity
Attitudes
Evaluation
Groups
Competition

ABSTRACT

In this research, we examine how sudden shifts in social identity can swiftly shape implicit evaluations. According to dual system models of attitudes, implicit attitude change is often slow and insensitive to explicit cues or goals. However, the social identity approach suggests that the intergroup context can shape nearly every aspect of social cognition from explicit preferences to implicit evaluations. In three experiments, we test whether explicit cues about social identity and the intergroup context can swiftly shape implicit evaluations. We find that people quickly develop an implicit preference favoring their in-group relative to the out-group—even when the group assignments are arbitrary. Importantly, this pattern of implicit intergroup bias quickly shifts following subtle changes in the intergroup context. When we frame the two groups as cooperative (vs. competitive), implicit intergroup bias is eliminated. Finally, being switched from one minimal group to the other reverses implicit intergroup bias, leading people to favor their new in-group (and former out-group). Individual differences in the degree to which people readily switch their implicit intergroup preference are correlated with their need to belong. In sum, these studies provide evidence that social identity cues and goals rapidly tune implicit evaluation. This research not only speaks to the influence of social identity on implicit cognition, but also has implications for models of attitude development and change.

People belong to multiple social groups—ranging from well-established groups such as race, gender, and nationality, to more fluid ones like schools, work groups, and sports teams. When group identities infuse our sense of self, people tend to perceive themselves and others as interchangeable exemplars of a social category as opposed to unique individuals (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). According to Turner, Oakes, Haslam, and McGarty (1994), the human capacity to adopt identities and using them as a lens to view the world implies that "All cognition is social cognition" (Turner et al., 1994, p. 462). Indeed, shifting from perceiving oneself and others as individuals to perceiving oneself and others as group members can exert a profound influence on how people perceive and interact with their social surroundings (Hastorf & Cantril, 1954; Postmes & Jetten, 2006; Turner et al., 1987; Xiao, Coppin, & Van Bavel, 2016). Yet, little work has examined how dynamic aspects of social identity and surrounding intergroup context shape implicit cognition—and particularly, the development *and change* of implicit attitudes.

One reason that social groups may exert a powerful impact on human cognition and behavior is that they help fulfill numerous critical social goals—from the need for distinctiveness (Brewer, 1991) to the need to belong (Baumeister & Leary, 1995). Theorists have argued that the need to belong is as central to psychological well-being as food and shelter are for physical well-being (Baumeister & Leary, 1995). For instance, people high in the need to belong are particularly accurate in decoding social cues (Pickett, Gardner, & Knowles, 2004), tend to have selectively superior memory for social events and information (Gardner, Pickett, & Brewer, 2000), and pay special attention to members of their in-group (Van Bavel, Swencionis, O'Connor, & Cunningham, 2012). Likewise, those whose belonging needs have been threatened (e.g., social exclusion or ostracism) tend to be more prosocial (Maner, DeWall, & Baumeister, 2007), contribute more on a group task (Williams & Sommer, 1997), and show greater recognition for positive emotions (Bernstein, Young, Brown, Sacco, & Claypool, 2008). In short, belonging needs dictate an extraordinary amount of human cognition and behavior.

People can fulfill their need to belong—as well as many other psychological needs—by forging interpersonal relationships and affiliating with social groups (Brewer, 1991). The functional link between basic social motives and group membership may help explain why group identities and discrimination can emerge rapidly under very trivial and

* Corresponding author at: School of Interdisciplinary Arts and Sciences, University of Washington, Tacoma, P.O. Box 358436, 1900 Commerce St, Tacoma, WA 98402, United States of America.
E-mail address: yxiao2@uw.edu (Y.J. Xiao).

https://doi.org/10.1016/j.jesp.2019.03.005
Received 12 October 2018; Received in revised form 2 March 2019; Accepted 4 March 2019
Available online 23 March 2019
0022-1031/

arbitrary circumstances (also see Dunham, 2018 for a recent review; Tajfel, Billig, Bundy, & Flament, 1971). For instance, in early investigations of the minimal group paradigm, people performed a trivial task such as guessing the number of dots in a rapidly presented image or expressing preference for abstract paintings from Klee and Kandinsky (Brown, Collins, & Schmidt, 1988; Tajfel et al., 1971). Surprisingly, even such minimal and arbitrary assignment of "groups" led people to express in-group favoritism in resource allocation, giving more money to anonymous in-group members, among other measures of in-group favoritism (Tajfel, 1982). These minimal group studies illustrate the ease with which people identify with and favor their in-group (Brewer, 1979). In the current research, we examine how these arbitrary social identities tune implicit intergroup preferences.

1. A social identity approach

There is now solid evidence that newly formed group identities can impact automatic perceptions and evaluations of the social world (Ashburn-Nardo, Voils, & Monteith, 2000; Otten & Moskowitz, 2000). For instance, merely assigning people to minimal mixed-race teams can eliminate implicit racial bias (Van Bavel & Cunningham, 2009) and alter responses in regions of the brain associated with automatic evaluation (Van Bavel, Packer, & Cunningham, 2008) and face processing (Ratner & Amodio, 2013; Van Bavel, Packer, & Cunningham, 2011). These findings suggest that activating a new identity—in this case, team membership—may temporarily override other forms of implicit bias (see also Scroggins, Mackie, Allen, & Sherman, 2016). Our work examines how features of the situation and individual differences shape the construction of these implicit evaluations.

The present research focuses specifically on implicit group evaluations towards *novel groups*, which is a fundamental component of implicit prejudice towards significant meaningful social groups (e.g., Brewer, 1979; Tajfel et al., 1971). Specifically, we study the role of social group affiliation in the formation, reduction, and reversal of implicit intergroup preferences. Understanding the relationship between social identity and implicit bias has become a source of considerable interest outside the academic community thanks to a growing industry focused on implicit bias training (Chabris & Brown, 2018), which has attracted skepticism (e.g., Jussim, 2017). The present research is designed to provide greater theoretical precision about the role of social identity in implicit evaluation, which may help understand why some attempts to reduce implicit bias fail and possibly lay the foundation for more effective strategies in reducing implicit biases (see Lai et al., 2016; Scroggins et al., 2016; Van Bavel & Cunningham, 2009). Because the need to belong is an important individual difference linked to identifying with social groups, we also test the possibility that this social need may predict the flexibility in people's implicit group preferences.

We examine whether people develop implicit intergroup bias when assigned to arbitrary groups, *and* if so, whether such implicit intergroup bias is driven by implicit in-group favoritism or implicit out-group derogation (Brewer, 1999). In studying intergroup biases, others have argued that more work should distinguish in-group favoritism versus out-group derogation (e.g., Hewstone, Rubin, & Willis, 2002; Brewer, 1979). However, such efforts have produced mixed findings regarding the processes underlying *implicit* intergroup biases. The mixed findings in the literature may be due to different types of groups studied (e.g., racial, ethnic, gender, or minimal groups) as well as different measures of implicit intergroup bias (e.g., IAT; affective priming). For instance, some of this previous work has found intergroup bias being driven by implicit in-group favoritism as opposed to out-group derogation (e.g. Otten & Moskowitz, 2000; Van Bavel & Cunningham, 2009), while others show a combination of both processes (see Nosek & Banaji, 2001 for a review). Meanwhile, some work has used tasks that make is difficult to dissociate implicit in-group favoritism from out-group derogation (e.g., Implicit Association Task; in Ashburn-Nardo, Voils, & Monteith, 2001). The current research will shed more light on inferences about the nature of implicit group preferences in minimal group scenarios.

2. Explicit and implicit attitudes

Attitudes have long been regarded as one of the most indispensable constructs in the history of social psychology (Allport, 1935; Eagly & Chaiken, 1993; Maio & Haddock, 2015; Petty & Cacioppo, 1981). Attitudes play a significant role in many social processes and behaviors, such as how we perceive ourselves and others (e.g., Dotsch, Wigboldus, & van Knippenberg, 2011; Young, Ratner, & Fazio, 2014), whether we intend to purchase certain products (e.g., Honkanen, Verplnken, & Olsen, 2006; Maison, Greenwald, & Bruin, 2004), and how we interact with certain social groups (e.g., Dovidio, Kawakami, & Gaertner, 2002; Penner et al., 2010). In recent decades, dual process models have served as the guiding theoretical paradigm in this domain, including topics such as self-control (e.g., Baumeister & Heatherton, 1996; Strack & Deutsch, 2004), stereotyping and prejudice (e.g., Devine, 1989; Greenwald & Banaji, 1995), persuasion (Petty & Cacioppo, 1986; Rydell & McConnell, 2006), and person perception (e.g., Brewer, 1988; Macrae & Bodenhausen, 2000).

Dual attitude theories have been proposed to explain two different forms of attitude – explicit and implicit attitude. Explicit attitudes generally refer to attitudes we can consciously report and control, and are often inferred from self-reported evaluative judgments (e.g., Bogardus, 1925; Thurstone, 1928). In contrast, implicit attitudes generally refer to attitudes that we are not consciously aware of and cannot consciously control, and are often inferred from indirect measures such as the Implicit Association Task (IAT; Greenwald, McGhee, & Schwartz, 1998) or sequential priming tasks (Fazio, Jackson, Dunton, & Williams, 1995). A growth in implicit measures has spawned countless studies on these differences over the past few decades and there has been a vigorous discussion about how these attitudes develop and change as well as their predictive validity (Fazio & Olson, 2003).

Explicit and implicit attitudes have been shown to independently predict behaviors (e.g., Dovidio et al., 2002) and choices (e.g., Galdi, Arcuri, & Gawronski, 2008). For instance, in one study, implicit and explicit racial attitudes predicted spontaneous and deliberate interracial behaviors, respectively (Dovidio et al., 2002). However, a recent meta-analysis suggests that implicit attitudes predict behaviors, even explicit ones, better than explicit attitudes do (Kurdi et al., 2018). Therefore, it is possible that implicit attitudes predict a wider range of behaviors more consistently than explicit attitudes, compared to what was previously believed. However, more work is needed to fully understand the relationship between implicit attitudes and behaviors.

3. Attitude change

Although attitude is often defined as "a general and enduring positive or negative feeling about some person, object or issue" (Petty & Cacioppo, 1981, p. 7), there is extensive evidence that our attitudes change. As such, many researchers have conducted research and proposed theoretical frameworks to understand and predict how implicit and explicit attitudes form and change (e.g., Cunningham, Zelazo, Packer, & Van Bavel, 2007; Gawronski & Bodenhausen, 2007; Greenwald & Banaji, 1995; Mann & Ferguson, 2015; Rydell & McConnell, 2006). Most dual attitude models share the assumption that implicit and explicit attitudes change through distinct processes, whereby implicit attitudes are more difficult and slower to change compared to explicit attitudes (e.g., Petty, Wegener, & Fabrigar, 1997; Rydell & McConnell, 2006; Smith & DeCoster, 2000).

Several theories of attitude change propose that implicit attitudes arise and change through the use of a slow-learning process and are affected by repeated associative pairings; whereas explicit attitudes arise and change through fast-learning, rule-based processes, and

affected by explicit processing goals (e.g., Gawronski & Bodenhausen, 2007; Rydell & McConnell, 2006; Sloman, 1996). Some theories have even argued that implicit attitudes can only shift slowly over time and are insensitive to single instances of new information contradicting prior learning, unaffected by explicit processing goals, and "exclusively affected by associative information about the attitude object that is not available for higher order cognition" (Rydell & McConnell, 2006, p. 995).

There has been extensive research demonstrating that implicit attitudes can form and change through associative learning (Olson & Fazio, 2001). For instance, repeated associative pairings such as subliminal and supraliminal evaluative priming, and exposure to repeated pieces of information about targets, tend to shape implicit attitudes (Gregg, Seibt, & Banaji, 2006; Rydell & McConnell, 2006). In contrast, explicit goals such as explicit instructions to rely or not rely on first impressions, tend to shape explicit attitudes as often revealed by self-report measures. These findings are consistent with theoretical frameworks with separate routes through which implicit versus explicit attitudes form and change (Gawronski & Strack, 2004; Gregg et al., 2006; Rydell & McConnell, 2006).

Similarly, studies have found that implicit attitudes towards social groups were "easier done than undone" (Gregg et al., 2006, p. 1). In order to examine attitude change, one set of studies examined two attitude change manipulations: "abstract supposition" was manipulated by informing participants that the novel groups they had previously learned about would switch characters, whereas "concrete learning" meant providing participants with a detailed and vivid narrative about how the two previously learned groups switched characters (Gregg et al., 2006). As a result, explicit attitudes towards these novel groups were reversed by both attitude change manipulations, but the *newly-formed* implicit attitudes were immune to such manipulations (Gregg et al., 2006). Therefore, it was concluded that implicit attitudes could be easily induced, but could not be shifted rapidly (Gregg et al., 2006). These findings suggest that changing *newly formed* implicit attitudes may require more time and exposure compared to developing these implicit attitudes in the first place. However, the groups participants learned about here did not implicate social identity and these findings may have therefore underestimated the potential capacity for people to shift their implicit attitudes along with their identity.

3.1. Flexibility of implicit attitudes

In contrast to earlier work, several recent studies have found that even implicit attitudes *can be* induced and shaped by explicit motivation or processing goals. For instance, randomly assigning people to groups has been shown to induce a preference for in-group over out-group members on measures of implicit evaluation (e.g., Ashburn-Nardo et al., 2001; Otten & Moskowitz, 2000; Van Bavel & Cunningham, 2009). Indeed, Blair's (2002) review of strategies to reduce implicit prejudice found that self and social motives were among the most effective ones for developing alternative implicit intergroup preferences (see also Ferguson & Bargh, 2004). Building on this work, our research examines whether explicit cues about social identity and the intergroup context can rapidly *alter* implicit evaluation—even undoing implicit preferences towards groups.

Along these lines, there is now evidence that implicit evaluations can be shifted, and even reversed, through providing single instances of new information that contradicts prior knowledge that was either deep-rooted (Van Dessel, Ye, & De Houwer, 2018) or just learned (Cone & Ferguson, 2015), or by offering reinterpretation of earlier information (Mann & Ferguson, 2015). In these studies, learning a new piece of information was able to change implicit evaluation of a well-known historic figure (Van Dessel et al., 2018) and implicit evaluations formed through repeated pairings (Cone & Ferguson, 2015). Moreover, the reinterpretation of previously learned information was also effective in shaping implicit evaluations, and such influence held after a 3-day delay (Mann & Ferguson, 2015). Similarly, in another study, when implicit attitude change occurred through reinterpretation of previously learned information or through learning new diagnostic information, such changes could generalize to novel contexts (Brannon & Gawronski, 2017). As such, it seemed plausible that shifts in identity might exert the same influence on implicit evaluations.

Other evidence suggests that certain language cues can rapidly alter implicit attitudes. For instance, recent research compared the effectiveness of repeated evaluative pairings (exposure to category members paired with pleasant or unpleasant stimuli) and evaluative statements (verbally signaling upcoming pairings without exposure; Kurdi & Banaji, 2017). Strikingly, one-shot language-based learning led to larger shifts in implicit attitudes than exposure to pairings that associate category members with pleasant or unpleasant images (Kurdi & Banaji, 2017). In another study, mere instructions about whether a stimulus would occur frequently or infrequently, in the absence of *actual* repeated exposure, was able to influence implicit stimulus evaluations (at least on some implicit measures; Van Dessel, Mertens, Smith, & De Houwer, 2017). These recent research findings offer evidence that implicit attitudes can be rapidly shifted—or undone—by explicit cues or motivation.

In this paper, we examine how identity shapes this formation of implicit intergroup preferences,[1] and how it can be rapidly changed by the group context. This topic is critical for understanding how to mitigate implicit bias, in hopes of potentially reducing subtle prejudicial behaviors and interactions (Blair, 2002). Until now, research studies trying to identify effective interventions to reduce implicit bias have yielded mixed findings (Blair, 2002; Devine, Forscher, Austin, & Cox, 2012; Hu et al., 2015; Lai et al., 2014, 2016). Many studies find that implicit intergroup biases—such as racial bias—are surprisingly resistant to change (Joy-Gaba & Nosek, 2010; Lai et al., 2016). Our work aims to address this issue by clarifying the role of social identity in these implicit social preferences.

4. Overview of current experiments

In the current research, we examine whether implicit evaluations can form and change quickly through shifts in social identity, rather than through slow associative processes as previously believed. We study this issue in the context of implicit group preferences, because we believe the motivation to belong to social groups can be strong enough to elicit changes in implicit evaluation. Moreover, since we do not require participants to know any individuating information about any group member, our research also provides unique evidence for the flexibility of implicit intergroup evaluations, extending previous research that replied on implicit associations formed about individuals (Cone & Ferguson, 2015; Van Dessel et al., 2018).

We focus on people's implicit preference towards newly introduced in-group and out-group for several reasons. First, we assess the rapid formation of implicit evaluations towards novel social categories. This aspect is important because whether newly formed implicit evaluations can readily shift has been a point of debate with mixed findings (Brannon & Gawronski, 2017; Cone & Ferguson, 2015; Gregg et al., 2006; Kurdi & Banaji, 2017; Mann & Ferguson, 2015; Van Dessel et al., 2017). Second, using novel groups allows us to manipulate the intergroup context in real time and measure corresponding shifts in implicit bias. This is important because we can then induce social motives among participants. Third, we believe that using minimal groups provides a stronger test of our research question, by observing the direct impact of social identity in the absence of real social connections to

[1] In this paper we do not distinguish between "implicit attitude" "implicit evaluation" and "implicit preference". Because we measure implicit attitude towards novel, newly introduced groups and individuals, these terms all refer to the same mental construct.

members from either group and avoiding the potential confound that any relationship exists between belonging needs and intergroup bias prior to the study. Fourth, our method examines the influence of very brief exposure of social identity cues (i.e., 30 ms), which is important for assessing *implicit* evaluation. Fifth, using minimal groups and assessing implicit group preferences using a priming procedure allow us to test the precise process underlying implicit intergroup preferences – whether they are driven by implicit in-group favoritism or out-group derogation. Sixth, by using priming rather than the Implicit Association Tests, we are able to examine the impact of identity cues in the absence of forced categorization (i.e., the IAT requires that participants categorize stimuli according to social categories which makes those categories salient during evaluation).

Overall, we test four main predictions. First, we predict that implicit group evaluations can be readily and quickly induced as a function of group assignment. This would conceptually replicate prior work establishing that implicit evaluations and attitudes can be formed quickly (e.g., Gregg et al., 2006)—especially findings from using minimal groups (Ashburn-Nardo et al., 2001; Scroggins et al., 2016; Van Bavel & Cunningham, 2009). Second, we predict that such implicit intergroup preferences can be highly attuned to the current intergroup context—leading to implicit intergroup bias during competition, but not cooperation. Third, we predict that previously formed implicit intergroup evaluation can be quickly shifted as soon as the group identities change. This happens every day when people move, change teams, or start a new job. Fourth, we examined the role in social motives—namely, the need to belong—on the flexibility to shift implicit intergroup preferences quickly. Thus, overall, we propose that implicit evaluations can be both induced *and* changed very swiftly according to social identity concerns.

We report three experiments examining how implicit intergroup evaluations can quickly form and change. In Experiment 1, we assign participants to minimal groups (red vs. blue team) and assess their implicit evaluations towards their own minimal in-group and out-group. If this replicates prior work, it would help validate a method for the subsequent experiments. In Experiment 2, we follow up by manipulating the intergroup context – informing participants whether the two groups were cooperative or competitive, in order to determine whether the formation of implicit group preferences is sensitive to changes in the intergroup context. In Experiment 3, we examine whether implicit group evaluations can change quickly *within* an individual, when their group assignment changes. Importantly, we also examine the extent to which people's own belonging needs play a role in the flexibility of shifting implicit intergroup preferences. While stable access to social groups and meaningful relationships is a universal human need (Baumeister & Leary, 1995), individual differences exist in the strength of people's belonging needs (Leary, Kelly, Cottrell, & Schreindorfer, 2013) and these differences are associated with a preference for in-group members (Van Bavel et al., 2012). These three experiments provide evidence for the role of social identification and belonging motives in shaping one's implicit group evaluations.

5. Experiment 1: group identities shape implicit evaluations

Our goal in Experiment 1 is to test the prediction that implicit group preferences can develop quickly when individuals are assigned to arbitrarily determined groups. In the interest of conceptually replicating previous research, we predict that participants will form relatively more positive implicit attitude towards their in-group compared to their out-group, as assessed by our response time measure. If our prediction is confirmed, it would be consistent with several other findings suggesting that implicit attitudes do not necessarily need to form from repeated associated pairings over time (Blair, 2002; Cone & Ferguson, 2015; Kurdi & Banaji, 2017; Mann & Ferguson, 2015; Van Dessel et al., 2017). In an era where numerous findings in the field have either failed to replicate or been called into question (OSC, 2015), we believe it is critical to first show that the effect of group identity on implicit evaluation is a robust phenomenon and then replicate and extend this method in each subsequent experiment. We believe this is an important finding for both the literature on implicit attitudes and intergroup relations.

5.1. Method

5.1.1. Participants

We recruited 65 undergraduate students (52.3% female and 46.2% female; gender information from one participant missing) at New York University (NYU) to participate in this study (named “Judge these pictures” for cover story) as partial fulfillment for course credit. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the set of experiments. For all the experiments reported in this paper, we collected data to satisfy at least 80% power assuming small to medium effect size for the main predicted interaction effects. Because experiment time slots were always posted a few weeks in advance and we could not perfectly predict how many participants would eventually sign up, the numbers fluctuate slightly from study to study. We also report sensitivity power analyses (in the Methods section) for each experiment. All data reported in this paper were collected from 2012 to 2014 and all the stimuli, code, and data will be made available upon publication at our OSF page.

5.1.2. Sensitivity power analysis

As recommended by the journal, for all experiments reported in this paper, we computed *sensitivity power analyses* using G*Power3 (Faul, Erdfelder, Lang, & Buchner, 2007) for the main predicted interaction effects. Because G*Power does not calculate interactions between repeated measures factors, we compute the sensitivity power analyses treating one of the repeated measures factor as a between groups factor (note that this is a more conservative power analysis compared to our repeated measures design). In Experiment 1, the minimum effect size that could be detected at 80% power (0.05 alpha level; average correlation coefficients among all repeated measures $r = 0.89$) for our main predicted 2-way interaction is $f = 0.08$. This indicates that our experiment and statistical test were sensitive to detect a small effect size (Cohen, 1988). G*power protocols for all of the power analyses reported in this paper are available at our OSF page.

5.1.3. Materials

5.1.3.1. Positive and negative target images. To use as our target pictures, we selected pictures from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008). IAPS was developed to provide a set of normative emotional stimuli, with normalized rating of various aspects of each picture, including valence. Target images were selected form IAPS on the basis of two requirements. First, images were clearly positive or negative, and second, they did not include emotional facial expressions. We selected 72 pictures with the highest ratings on valence (half positive and half negative), because we needed unambiguously positive and negative pictures as target stimuli in our experiment. The positive images included flowers, kittens, bunnies, and nature scenes, etc. The negative images selected included cockroaches, spiders, snakes, car accidents, and open wounds, etc. On a 9-point scale, the 36 positive images we used had valence ratings of $M = 3.30$ ($SD = 0.63$), and the 36 positive images we used had valence ratings of $M = 7.31$ ($SD = 0.51$), $t(70) = -29.78$, $p < .001$, $d = 7.02$, with higher numbers indicating greater positivity (valence rating obtained from Lang et al., 2008). The list of images is available at our OSF page.

5.1.3.2. Facial photographs. We used photos from Radbound Faces Database (RaFD; Langner et al., 2010) as ostensible group members during the learning phase of the study. These face images displayed neutral facial expressions. We used photos of a total of 36 individuals, thus 18 members in each group. There was an even split between male

and female individuals within each group. The photos were counterbalanced between the in-group and out-group across participants to ensure that any difference between the photos could not account for any group differences.

5.1.3.3. Collective identification scale. We assessed participants' level of collective identification with their respective minimal in-group, with a 3-item collective identification measure (e.g., I do not value being a member of the RED team – reverse coded; Van Bavel & Cunningham, 2012). One of the three items on the scale is reversed coded. Participants responded on a 7-point Likert scale (*1 = strongly disagree* to *7 = strongly agree; M* = 3.38, *SD* = 1.15, Cronbach's α = 0.65) and the results suggested that participants had a moderate level of identification with the in-group. The order in which the items appeared was completely randomized among participants (please see Appendix A for the complete scale).

5.1.3.4. Need to Belong (NTB) scale[2]. We measured participants' belonging needs using the 10-item Need to Belong scale (Baumeister & Leary, 1995). Sample items include "I want other people to accept me", "I seldom worry about whether other people care about me" (reverse coded). Half of the items on the scale are reversed coded. Participants responded on a 7-point Likert scale (*1 = strongly disagree to 7 = strongly agree;* M = 4.54, SD = 0.86; Cronbach's α = 0.79) with results suggesting that participants had a moderately high need to belong. The order in which the items appeared was completely randomized among participants (please see Appendix B for the complete scale).

5.1.3.5. Political orientation[3]. Then we assessed participants' political orientation with a 1-item measure: Where on the following scale of political orientation would you place yourself (overall, in general)? Participants responded on a 7-point scale (*1 = Extremely liberal; 7 = Extremely conservative*).

5.1.4. Procedure

5.1.4.1. General instructions and learning phase. After completing informed consent, participants followed instructions on a Dell computer screen. First participants were told that we were interested in better understanding the cognition behind the way people categorize images, and in this study they would categorize images based on their valence (i.e., positive vs. negative). Then we told participants that they would be assigned to one of two teams (RED or BLUE), and they would see picture of people on their own team and the other team. To indicate each person's group membership, we had face photos on a red background or a blue background.

Participants then went through a learning phase in which they were instructed to try their best to remember faces of individuals from the two groups. We presented 18 faces in blocks by group membership. Within each block, each face was presented for 6 s (adapted from Brosch & Van Bavel, 2012). There was a fixation cross that appeared in the center of the screen for 1 s before each face appeared. Each block was presented twice to facilitate memory. The order in which the blocks appeared was randomized. The order in which faces appeared within each block was also randomized.

5.1.4.2. Test phase. After the learning phase, participants were given instructions of the actual trials and were given a chance to practice. On each trial, participants were asked to focus their attention on the fixation cross that appears at the middle of the screen for 1 h. Subsequently, a red or blue color image chosen at random appeared on screen for 30 ms immediately before a target (IAPS) picture appeared. At this point, participants were instructed to indicate as fast and accurately as possible whether this target image was positive or negative in valence (i.e., "good" or "bad") by pressing "1" or "0" keys (for similar procedures, see Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Lowery, Hardin, & Sinclair, 2001; Weisbuch & Ambady, 2008). After they made a response, there was an inter-trial interval of 2 h before the next trial began (please see Fig. 1 for a representation of the task). Response accuracy and reaction times were recorded via *E*-prime software.

Participants were given the opportunity to go through two practice trials and a chance to ask any question before starting the actual test phase. Participants saw the red group prime on one practice trial and the blue group prime on the other practice trial. The two target IAPS images that appear on the practice trials do not appear in any of the actual trials in the test phase.

In the test phase, there were four blocks of trials in this experiment, with an opportunity to take a short break in between blocks. Each block contained 72 trials, and there were 288 trials in the entire experiment. On each trial, the pairing between the group symbol color photo and target image (IAPS) was chosen completely at random without replacement within each block.

5.1.4.3. Questionnaires. After participants finish the last block of trials, they were tested on their memory of the previously seen faces, to be consistent with our cover story. Participants saw all 36 images (from both teams) and indicated which team (red or blue) each individual belonged to. Finally, we measured participants' collective identification with their respective in-group, need to belong, political ideology, race, and gender. Then all participants were debriefed and thanked for their participation.

5.2. Results

5.2.1. Data reduction

Prior to analysis, we eliminated reactions times on trials with incorrect responses (e.g., responding "good" when the target image was negative). Additionally, we eliminated trials with reaction times less than 200 ms or more than 3 SDs above each participant's mean RT (e.g., Weisbuch & Ambady, 2008).[4] We then log transformed all reaction times. All analyses were conducted on the log transformed RTs, but figures are made with actual RTs for ease of interpretation. These data reduction methods were held constant across all experiments reported in this paper.

[2] We included the Need to Belong (NTB) scale in all three experiments reported here. However, Experiments 1 and the competitive condition in Experiment 2 were underpowered to examine the 3-way interactions. Sensitivity power analyses indicated that the sample sizes would only be sensitive to detect a moderate to large effect size to assess the role of NTB as a moderator of implicit intergroup preference Experiment 1 and the competitive condition in Experiment 2. Therefore, NTB was only analyzed in Experiment 3 in a correlation analysis (see Experiment 3 for more detail).

[3] We included the one-item political orientation measure to all the studies we ran in our lab during this period, to separately examine correlates of political orientation for a different project. Therefore, this item is reported here but not analyzed for the current experiments. The data will be made available for anyone who wishes to analyze this item.

[4] To our knowledge, difference researchers and labs tend to choose slightly different cut-offs for reaction time data reduction. For instance, some eliminated reaction times (RTs) greater than three standard deviations above each individual's mean RT, on an affective priming task (Lowery et al., 2001; Weisbuch & Ambady, 2008). Others have eliminated RTs below 300 ms and above 3000 ms (Dasgupta & Greenwald, 2001), or eliminated those below 150 ms and above 5000 ms (Gregg et al., 2006), on an Implicit Association Task (IAT) – another widely used reaction time task measuring implicit evaluation. Because our task is most similar to those used in Weisbuch and Ambady (2008) and Lowery et al. (2001), we used 3SD as our upper cut-off, and added a lower cut-off of 200 ms.

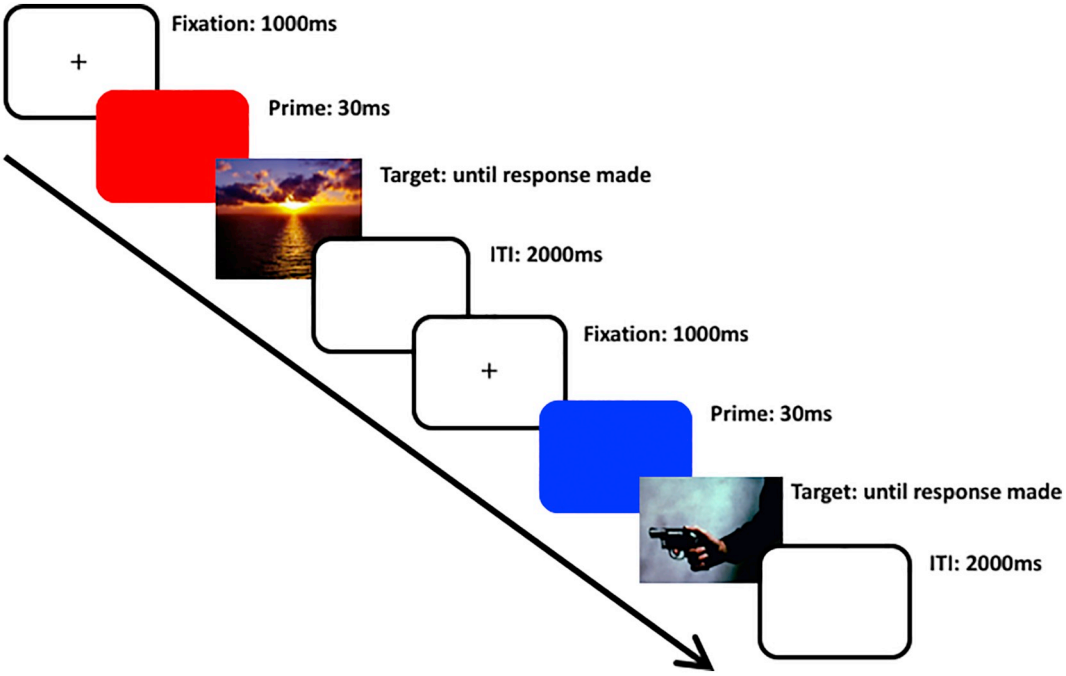

Fig. 1. Representation of two trials in the sequential priming task used in all experiments reported in this paper. Each trial starts with a fixation cross at the center of the screen for 1 s, followed by the prime (red or blue) for 30 ms. Then the target image appears immediately and stays on the screen until a response is made, followed by an inter-trial interval (ITI) for 2 s before the next trial starts. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2.2. Reaction times

In this experiment, we predicted that participants would develop implicit preferences (as revealed by reaction times in our task) towards their respective in-group compared to their respective out-group. To test our hypothesis, we subjected log transformed RTs to a 2 (prime identity: in-group vs. out-group) x 2 (target valence: negative vs. positive) repeated analysis of variance (ANOVA). The analysis revealed a significant main effect of prime identity, F (1, 64) = 12.32, p = .001, η^2 = 0.16, such that overall participants responded to target pictures faster after exposure to an out-group face, and a main effect of target valence, F (1, 64) = 7.28, p = .009, η^2 = 0.10, such that participants responded to negative target pictures faster. These effects were qualified by a significant prime identity x target valence interaction, F (1, 64) = 17.51, p < .001, η^2 = 0.22, indicating intergroup bias such that that participants' response times to valenced pictures were influenced by the group identity of person in the prime photo (see Fig. 2 & Table 1). Specifically, simple effects analysis revealed that participants responded faster to negative images compared to positive images after exposure to the out-group color symbol, F (1, 64) = 18.73, p < .001, η^2 = 0.23. RTs to positive and negative images did not differ significantly after exposure to in-group color symbols, F (1, 64) = 0.37, p = .54, η^2 = 0.01.[5] Thus, the implicit intergroup preference was consistent with out-group derogation in this study.

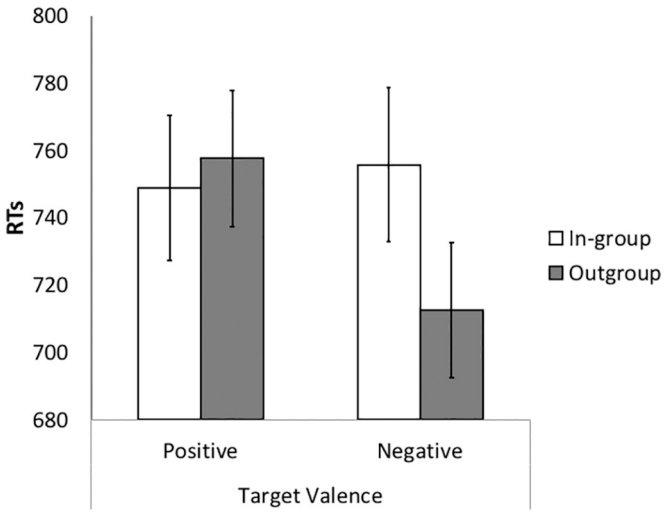

Fig. 2. Raw reaction times (RT; on Y-axis) are plotted as a function of prime identity (in-group vs. out-group) and target valence (positive vs. negative). Raw reaction times are presented in all figures for ease of interpretation, and when analyzed, do not change pattern of results. There is a significant two-way interaction between prime identity and target valence indicating intergroup bias. Error bars represent standard errors of the mean.

5.3. Discussions

In Experiment 1, we replicated previous research finding that implicit group attitudes can form quickly once group memberships are established. Using the minimal group paradigm, we assigned participants to one of two groups based on rather arbitrary basis (i.e., color) then assessed their implicit evaluation of individuals that belong to the same group, and individuals of an out-group, using a modified affective priming task. Conceptually replicating findings from prior research (Ashburn-Nardo et al., 2001; Otten & Moskowitz, 2000; Van Bavel & Cunningham, 2009), participants formed implicit attitudes readily and immediately after being assigned to a group. Specifically, participants responded faster to negative images compared to positive images after exposure to the out-group color symbol, indicating that the implicit intergroup bias we observe here appeared to be driven by out-group derogation.

[5] Participants responded faster to negative images after exposure to out-group color symbols, compared to after exposure to in-group color symbols, F (1, 64) = 35.48, p < .001, η^2 = 0.36. RTs to positive images did not differ significantly after exposure to out-group versus in-group color symbols, F (1, 64) = 1.67, p = .20, η^2 = 0.03.

Table 1
Mean response times (RTs) and standard deviations (Exp 1).

	Target valence					
	Positive			Negative		
	M (SD)	95% CI		M (SD)	95% CI	
		Lower	Upper		Lower	Upper
In-group	748.91 (173.09)	706.02	791.80	755.56 (183.66)	710.22	801.24
Out-group	757.58 (162.85)	717.23	797.93	712.56 (162.12)	672.39	752.73

Note: Experiment 1 Means and Standard Deviations of raw reaction times (RTs) for all four cells: 2 (in-group vs. out-group) x 2 (target valence: positive vs. negative) (N = 65). M = Means, SD = Standard Deviation, CI = Confidence Interval.

In Experiment 2, we extend these findings by examining whether the quick formation of implicit group preferences could be sensitive to shifts in intergroup context. Specifically, we manipulate the minimal groups to be either competitive or cooperative with each other, and examine group members' implicit evaluations towards the in-group and the out-group. To our knowledge, no previous research has examined how the immediate context shapes these basic identity-driven implicit evaluations.

6. Experiment 2: cooperative vs. competitive contexts shape implicit evaluation

In Experiment 2, we test whether the formation of implicit intergroup preference can be flexibly shaped by the current intergroup context. We predict that implicit intergroup preferences form quickly as a function of group assignment, and these implicit evaluations are highly sensitive to the current intergroup context. Specifically, we randomly assign participants to a competitive or cooperative intergroup context. There is extensive evidence that intergroup dynamics are dramatically different when the context is competitive versus cooperative (e.g., Cikara & Van Bavel, 2014). As such, we expect implicit intergroup bias to disappear among participants who learned that the in-group and out-group are in a cooperative environment.

As we noted above, some previous research has argued that implicit attitudes are “easier done than undone” (Gregg et al., 2006, p. 1). Thus, it is possible that social identities might be instrumental in creating implicit preferences but that implicit preferences might be hard to change or eradicate once they are created. If this is the case, participants in this experiment should form more positive implicit evaluations towards the in-group than the out-group, and this pattern should not be dependent on the intergroup context.

To generate the strongest possible test of implicit flexibility, we make participants aware of the intergroup context *after* they have been assigned to groups and *immediately before* the implicit evaluation measurement. Hiding the context from participants until this phase of the experiment helps rule out the alternative explanation that participants have some time (although only a few minutes) to repeatedly process this information and build new associations. We work to minimize any such psychological process and test the true limits of rapid shifts in implicit evaluation.

6.1. Method

6.1.1. Participants

We recruited 100 undergraduate students (32% male and 68% female) at New York University (NYU) to participate in this study (“Judge these pictures”) as partial fulfillment for course credit (note that we recruited many more participants for this study since it was a mixed-design with a between-subjects condition).

6.1.2. Sensitivity power analyses

As recommended by the journal, for all experiments reported in this paper, we computed *sensitivity power analyses* using G*Power3 (Faul et al., 2007) for the main predicted interaction effects. In Experiment 2, we first calculated partial correlations between all the repeated measures controlling for the between-groups factor (competitive vs. cooperative). The minimum effect size that could be detected at 80% power (0.05 alpha level; average partial correlation coefficients among all repeated measures r = 0.91) for our main predicted 3-way interaction is f = 0.05. This indicates that our experiment and statistical test were sensitive to detect a small effect size.

6.1.3. Materials

All research materials were consistent with those from Experiment 1. Similarly, the 3-item collective identification scale was administered towards participants' respective in-groups (M = 3.30, SD = 1.27, Cronbach's α = 0.68) suggesting that participants had a moderate level of identification with the in-group. The 10-item Need to Belong scale was administered (M = 4.43, SD = 0.95, Cronbach's α = 0.78) suggesting that participants has moderately high level of need to belong.

6.1.4. Procedure

We followed exactly the same procedure as Experiment 1 with only one change. Participants were randomly assigned to the *competitive* or *cooperative* contexts (between subjects). Immediately before participants started the actual trials of the priming task assessing their implicit evaluations, and after they went through the practice trials, we told them: “Please keep in mind that the BLUE and RED teams are competing against (cooperating with) each other! As a member of the RED/BLUE team, you will be in competition with (cooperating with) the other team (BLUE/RED) later on in this study. Please keep this in mind during this study”. Participants' mean level of collective in-group identification did not significantly differ between the competitive (M = 3.36, SD = 1.21) and cooperative (M = 3.22, SD = 1.34) conditions, t (97) = 0.55, p = .58, d = 0.11.

6.2. Results

6.2.1. Data reduction

Data reduction was done in the same way as in Experiment 1.

6.2.2. Reaction times

We subjected log transformed RTs to a 2 (prime identity: in-group vs. out-group) x 2 (target valence: negative vs. positive) x 2 (context: competition vs. cooperation) mixed analysis of variance (ANOVA). The analysis revealed a significant main effect of prime identity, F (1, 98) = 25.97, p < .001, η^2 = 0.21, such that overall participants responded to target pictures faster after exposure to an out-group face, and a main effect of target valence, F (1, 98) = 7.65, p = .007, η^2 = 0.07, such that overall participants responded to negative target pictures faster. There was also a main effect of context, F (1, 98) = 6.34, p = .013, η^2 = 0.06, such that overall participants in the competitive context responded faster compared to those in the cooperative context. Replicating Study 1, we also found a significant prime identity x target

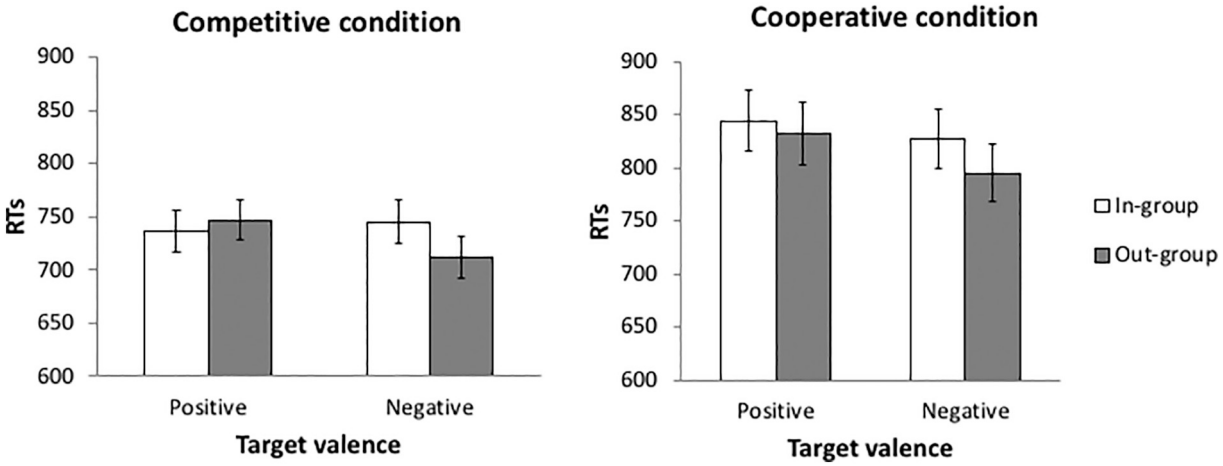

Fig. 3. Raw reaction times (RTs; on Y-axis) are plotted as a function of prime identity (in-group vs. out-group) and target valence (positive vs. negative) for participants in the competitive and the cooperative contexts. There is a marginally significant three-way interaction: prime identity x target valence x intergroup context. Error bars represent standard errors of the mean.

valence interaction, F (1, 98) = 15.67, $p < .001$, $\eta^2 = 0.138$, indicating that participants' response times to positive and negative pictures were influenced by the identity of the prime.

Importantly, we predicted that this quick formation of implicit group preferences would be sensitive to the manipulation of intergroup context. Consistent with our prediction, this prime identity x target valence two-way interaction was further qualified by a marginally significant 3-way interaction, F (1, 98) = 3.80, $p = .054$, $\eta^2 = 0.04$. Given this 3-way interaction, we conducted analyses separately for the competitive and cooperative contexts to unpack these findings.

6.2.2.1. Competitive context. We subjected log transformed RTs from the competitive context to a 2 (prime identity: in-group vs. out-group) x 2 (target valence: negative vs. positive) repeated analysis of variance (ANOVA). The analysis revealed a significant main effect of prime identity, F (1, 52) = 7.89, $p = .007$, $\eta^2 = 0.13$, and a marginally significant main effect of target valence, F (1, 52) = 3.83, $p = .056$, $\eta^2 = 0.07$. These were qualified by a significant 2-way interaction, F (1, 52) = 23.10, $p < .001$ (see Fig. 3 & Table 2). This interaction pattern replicated results from Experiment 1, indicating that participants had relatively more positive implicit evaluation of in-group symbols than out-group symbols.

Specifically, simple effects analysis revealed that participants responded faster to negative images compared to positive images after exposure to the out-group color symbol, F (1, 52) = 14.20, $p < .001$, $\eta^2 = 0.22$. RTs to positive and negative images did not differ significantly after exposure to in-group color symbols, F (1, 52) = 1.25, $p = .27$, $\eta^2 = 0.02$.[6] Thus, the implicit intergroup preference was consistent with out-group derogation in this study, consistent with findings from Experiment 1.

6.2.2.2. Cooperative context. We subjected log transformed RTs from the cooperative context to a 2 (prime identity: in-group vs. out-group) x 2 (target valence: negative vs. positive) repeated analysis of variance (ANOVA). The analysis revealed a significant main effect of prime identity, F (1, 46) = 17.63, $p < .001$, $\eta^2 = 0.28$, and a marginally significant main effect of target valence, F (1, 46) = 3.84, $p = .056$, $\eta^2 = 0.08$. Unlike the competitive context, the 2-way interaction was not statistically significant, F (1, 46) = 1.56, $p = .218$, $\eta^2 = 0.03$ (see Table 3). Therefore, no simple effect analysis was conducted for the Cooperative context. The pattern of implicit intergroup bias towards in-group and out-group members was not present in the cooperative context, suggesting that manipulating intergroup context shaped implicit attitudes.

6.3. Discussion

In Experiment 2, we replicated and extended the findings from Experiment 1. Specifically, we found evidence that recently formed implicit evaluations are sensitive to subtle cues in the intergroup context. When the two groups were presented as competitive, our findings were highly consistent with the patterns found in Experiment 1, such that participants developed more positive implicit evaluations towards the minimal in-group they were assigned to, compared to the minimal out-group. This could imply that the competitive mindset resembles the default mindset when people are assigned to groups, even when such group assignment is arbitrary. Also consistent with Experiment 1, implicit intergroup preference in the competitive condition was mostly driven by implicit derogation of the out-group.

However, when the two groups were presented as cooperative – thus reducing the need for intergroup distinctions and possible animosity—this pattern of implicit preference was no longer significant. Interestingly, participants in the competitive context responded faster on average compared to participants in the cooperative context. This may be due to the possibility that the competitive intergroup context induces a mindset that enhances engagement, and resembles a default mindset when people are assigned to groups, consistent with the patterns of implicit intergroup bias.

Therefore, not only can implicit intergroup preferences form as a result of social group identification, they are also highly sensitive to the relevant contextual information. In this way, it appears that implicit evaluations can be easily done *and* undone. However, in Experiment 2, the effect of context was a *between-* rather than *within*-participant factor, making it difficult to infer if these implicit evaluations would be actually "undone" within a single person. We addressed this issue in the next experiment.

7. Experiment 3: trading teams changes implicit evaluations

In Experiment 3, we move on to test whether implicit evaluations could be reversed quickly once they are formed. In order to test this question in an intergroup context, we employ a group switch procedure

[6] Participants responded faster to negative images after exposure to out-group color symbols, compared to after exposure to in-group color symbols, F (1, 52) = 32.64, $p < .001$, $\eta^2 = 0.39$. RTs to positive images did not differ significantly after exposure to out-group versus in-group color symbols, F (1, 52) = 3.07, $p = .09$, $\eta^2 = 0.06$.

Table 2
Mean response times (RTs) and standard deviations (Exp 2 competitive condition).

	Target valence					
	Positive			Negative		
	M (SD)	95% CI		M (SD)	95% CI	
		Lower	Upper		Lower	Upper
In-group	736.18 (141.37)	697.22	775.15	745.54 (147.00)	705.02	786.06
Out-group	746.88 (139.35)	708.47	785.29	712.30 (140.60)	673.55	751.06

Note: Experiment 2 Means and Standard Deviations of raw reaction times (RTs) for all four cells: 2 (in-group vs. out-group) x 2 (target valence: positive vs. negative) for the competitive context (N = 53). M = Means, SD = Standard Deviation, CI = Confidence Interval.

Table 3
Mean response times (RTs) and standard deviations (Exp 2 cooperative condition).

	Target valence					
	Positive			Negative		
	M (SD)	95% CI		M (SD)	95% CI	
		Lower	Upper		Lower	Upper
In-group	843.82 (197.48)	785.84	901.80	827.72 (190.54)	771.78	883.67
Out-group	831.53 (201.99)	772.22	890.83	795.17 (181.53)	741.87	848.46

Note: Experiment 2 Means and Standard Deviations of raw reaction times (RTs) for all four cells: 2 (in-group vs. out-group) x 2 (target valence: positive vs. negative) for the cooperative context (N = 47). M = Means, SD = Standard Deviation, CI = Confidence Interval.

in which participants are first assigned to a minimal group (as they were in previous experiments), and then switched to the opposite group halfway through the experiment. This is a common practice in sports, where players are regularly traded to another team during the season. We assess their implicit group evaluations both before and *immediately* after the group switch. We predict that when the relevant group contingencies and identities change, the previously formed implicit group evaluations will shift accordingly, and will do so quickly.

As a secondary hypothesis, in Experiment 3, we directly measure each participant's belonging needs, and assess whether this individual difference predicts *evaluative flexibility*. Evaluative flexibility represents the extent to which a person readily and quickly shifts their implicit evaluation. In the context of the current experiment, evaluative flexibility specifically refers to the extent to which participants shift their implicit preference towards the *current* in-group, after being switched to a new group. We predict that the higher an individual's belonging needs, the more readily they would reverse their implicit group preferences. If so, this would provide evidence that shifts in implicit group preferences can be predicted by social motives.

7.1. Method

7.1.1. Participants

We recruited 53 undergraduate students (22.6% male and 75.5% female; one missing gender information) at New York University (NYU) to participant in this study ("Judge these pictures") as partial fulfillment for course credit.

7.1.2. Sensitivity power analyses

As recommended by the journal, for all experiments reported in this paper, we computed *sensitivity power analyses* using G*Power3 (Faul et al., 2007) for the main predicted interaction effects. Because G*Power does not calculate interactions between repeated measures factors, we compute the sensitivity power analyses treating one of the repeated measures factors as between groups factors. Note that this is a more conservative power analysis compared to our repeated measures design. In Experiment 3, the minimum effect size that could be detected at 80% power (0.05 alpha level; average correlation coefficient among all repeated measures r = 0.77) for our main predicted 3-way interaction is f = 0.11. This indicates that our experiment and statistical test were sensitive to detect a small effect size.

The minimum effect size that could be detected at 80% power (0.05 alpha level) for the correlation between Evaluative Flexibility and Need to Belong is r = 0.37. This indicates that our experiment and statistical test were sensitive to detect a moderate correlation (Cohen, 1988).

7.1.3. Materials

All research materials were consistent with those from Experiments 1 and 2. Prior to the group switch manipulation, the 3-item collective identification scale was administered towards participants' original in-groups (M = 3.53, SD = 1.08, Cronbach's α = 0.89). After the group switch manipulation, the 3-item collective identification scale was administered towards participants' new in-groups (M = 3.36, SD = 1.09, Cronbach's α = 0.76). Participants had moderate levels of collective identification with the original in-group before group switch, *and* with the new in-group after the group switch.

The 10-item Need to Belong scale was administered at the end of the experiment (M = 4.68, SD = 1.01; Cronbach's α = 0.87) suggesting that participants has moderately high level of need to belong.

7.1.4. Procedure

We followed the same procedure as Experiment 1 with two changes. First, after participants finished two (of ostensibly four) blocks of trials, the experimenter told them that there was a computer error and they were accidentally assigned to the wrong team. Then participants were told that correct team assignment was important for our experiment, and we had to re-start them for the study with the correct team assignment. We also assured participants that we would only have them complete another two blocks of trials, and they would get out of the experiment in about the same amount of time. Then the experimenter went through a few screens of instructions with the participants to ostensibly make sure that they indeed received the correct team assignment. In fact, every participant was assigned to the other team (original out-group) after this manipulation, making this a within-

subjects design. This served as our group switch manipulation. In this study, participants finished all the trials, they filled out the Need to Belong scale, political orientation item, gender, and race.

7.2. Results

7.2.1. Data reduction

Data reduction was done in the same way as in Experiments 1 and 2.

7.2.2. Reaction times

Because of the nature of our experiment design, when we report our analyses here, we always use "in-group" to refer to the group each participant was assigned to during the first part of the experiment (i.e., *original* in-group), which became their out-group in the second part of the experiment after group switch. Similarly, we always use "out-group" to refer to each participant's out-group during the first part of the experiment before group switch (i.e., *original* out-group), would became their in-group in the second part of the experiment.

We subjected log transformed RTs to a 2 (prime identity: original in-group vs. original out-group) x 2 (target valence: positive vs. negative) x 2 (group switch: before vs. after) mixed analysis of variance (ANOVA). The analysis revealed a significant main effect of group switch, F (1, 52) = 30.20, $p < .001$, $\eta^2 = 0.37$, such that overall participants responded to target pictures faster after group switch. We also found a significant prime identity x group switch interaction, F (1, 52) = 12.83, $p = .001$, $\eta^2 = 0.20$, indicating participants' response time to original in-group vs. original out-group members differed before vs. after group switch. Importantly, these effects were qualified by a significant 3-way interaction, F (1, 52) = 10.88, $p = .002$, $\eta^2 = 0.17$. Therefore, we conducted analyses separately for before vs. after group switch to decompose down these effects.

7.2.2.1. Before group switch. We subjected log transformed RTs before group switch to a 2 (prime identity: original in-group vs. original out-group) x 2 (target valence: negative vs. positive) repeated analysis of variance (ANOVA). The analysis revealed a significant main effect of prime identity, F (1, 52) = 9.00, $p = .004$, $\eta^2 = 0.15$, indicating that participants responded faster to out-group members compared to in-group members. Importantly, this was qualified by a significant 2-way interaction, F (1, 52) = 8.66, $p = .005$, $\eta^2 = 0.14$, indicating that response time to positive vs. negative images depended on the original group membership of the person in the prime that appeared before the target images (see Fig. 4 & Table 4). This pattern was similar to results from Experiment 1 and the competitive context in Experiment 2, indicating that participants had relatively more positive implicit evaluation of in-group symbols, compared to out-group symbols, after a brief 30 ms exposure to them.

Simple effects analysis revealed that, prior to group switch, participants responded faster to negative images compared to positive images after exposure to the out-group color symbol, F (1, 52) = 9.52, $p = .003$, $\eta^2 = 0.16$. RTs to positive and negative images did not differ significantly after exposure to in-group color symbols, F (1, 52) = 0.24, $p = .63$, $\eta^2 = 0.01$.[7] Thus, the implicit intergroup preference prior to group switch was consistent with out-group bias in this study, consistent with findings from Experiment 1 & Experiment 2 (in the competitive context).

7.2.2.2. After group switch. We subjected log transformed RTs from after the group switch manipulation to a 2 (prime identity: in-group vs. out-group) x 2 (target valence: negative vs. positive) repeated analysis of variance (ANOVA). The analysis revealed a significant main effect of prime identity F (1, 52) = 6.06, $p = .017$, $\eta^2 = 0.02$, indicating that participants responded to the original in-group (now out-group) members faster. This was qualified by a marginally significant 2-way interaction, F (1, 52) = 3.21, $p = .079$, $\eta^2 = 0.06$, such that response time to positive vs. negative images depended on the group membership of the person in the prime photo that appeared before the target image. Importantly, this pattern of interaction was conceptually similar to the pattern from before group switch, indicating that the pattern of implicit preferences *reversed* after group switch. In other words, participants showed a relatively more positive implicit evaluation towards the original out-group (now the in-group) compared to the original in-group (now the out-group) (see Fig. 5 & Table 5 for raw RT data).

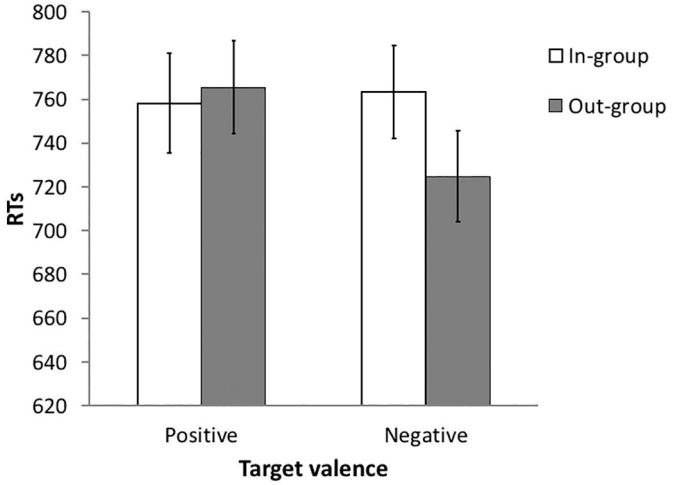

Fig. 4. Raw reaction times (RTs; on Y-axis) are plotted as a function of prime identity (in-group vs. out-group) and target valence (positive vs. negative) for participants prior to the group switch manipulation. There is a significant two-way interaction: prime identity x target valence, indicating an implicit preference towards the in-group. Error bars represent standard errors of the mean (N = 53).

Simple effects analysis revealed that, after group switch, reaction times to positive and negative images did not differ significantly after exposure to the *original* in-group (now out-group) color symbol, F (1, 52) = 2.12, $p = .15$, $\eta^2 = 0.04$. RTs to positive and negative images also did not differ significantly after exposure to in-group color symbols, F (1, 52) = 0.28, $p = .60$, $\eta^2 = 0.005$.[8]

7.2.2.3. Evaluative flexibility & Need to Belong (NTB). Following the calculation of "affective scores" from previous research using similar affective priming tasks (e.g., Lowery et al., 2001; Sinclair, Lowery, Hardin, & Colangelo, 2005), we calculated an Evaluative Flexibility index to reflect the extent to which each participant could flexibly reverse their implicit preference towards the group they currently belong to, after our group switch manipulation.

Specifically, we first subtracted average response times to positive images from average response times to negative targets for each prime category (i.e., "original in-group before group switch", "original out-group before group switch", "original in-group after group switch", and "original out-group after group switch"). These scores essentially represent implicit evaluation towards each prime category: positive

[7] Participants responded faster to negative images after exposure to out-group color symbols, compared to after exposure to in-group color symbols, F (1, 52) = 14.37, $p < .001$, $\eta^2 = 0.22$. RTs to positive images did not differ significantly after exposure to out-group versus in-group color symbols, F (1, 52) = 1.21, $p = .28$, $\eta^2 = 0.02$.

[8] Participants responded faster to negative images after exposure to original in-group color symbols, compared to after exposure to original out-group color symbols, F (1, 52) = 8.60, $p = .005$, $\eta^2 = 0.14$. RTs to positive images did not differ significantly after exposure to original out-group versus original in-group color symbols, F (1, 52) = 0.03, $p = .86$, $\eta^2 = 0.001$.

Table 4
Mean response times (RTs) and standard deviations (Exp 3 before group switch).

	Target valence					
	Positive			Negative		
	M (SD)	95% CI		M (SD)	95% CI	
		Lower	Upper		Lower	Upper
Original in-group	758.10 (166.05)	712.33	803.87	763.16 (154.81)	720.49	805.83
Original out-group	765.46 (155.33)	722.65	808.28	724.84 (150.11)	683.47	766.22

Note: Experiment 3 Means and Standard Deviations of raw reaction times (RTs) for all four cells: 2 (in-group vs. out-group) x 2 (target valence: positive vs. negative) prior to group switch (N = 53). M = Means, SD = Standard Deviation, CI = Confidence Interval.

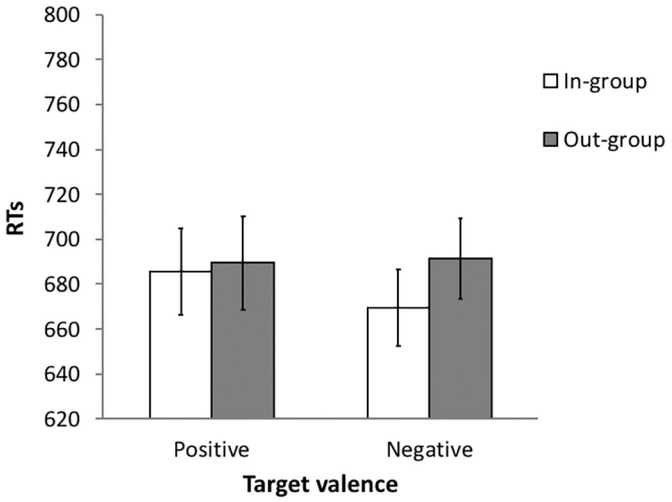

Fig. 5. Raw reaction times (RTs; on Y-axis) are plotted as a function of the *original* prime identity (*original* in-group vs. *original* out-group) and target valence (positive vs. negative) for participants after the group switch manipulation. There is a significant two-way interaction: prime identity x target valence. This indicates a reversal of implicit intergroup preferences after switching groups, such that participants showed an implicit preference towards the *original* out-group (now in-group). Error bars represent standard errors of the mean (N = 53).

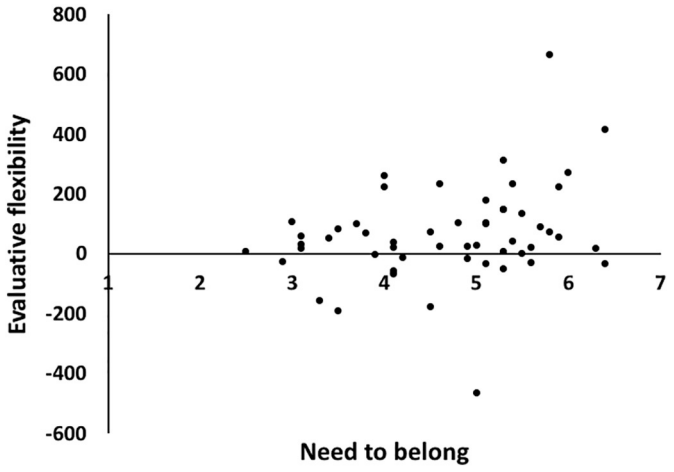

Fig. 6. A moderate positive correlations between individual Need to Belong scores (x-axis) and Evaluative Flexibility scores (y-axis; see explanation of calculation in-text): r (53) = 0.29, p = .035. Higher scores on the x-axis reflect a greater need to belong; higher scores on the y-axis reflect a stronger shift in implicit preference as a function of the *current* group assignment.

scores should be indicative of relatively positive evaluation and negative scores indicative of relatively negative evaluation (Lowery et al., 2001; Sinclair et al., 2005). Then, each participants' Evaluative Flexibility (EF) score is calculate as: **EF = (original in-group affect before group switch – original out-group affect before group switch) – (original in-group affect after group switch – original out-group affect after group switch)**. Thus, higher scores are indicative of relatively stronger shift in implicit preference as a function of the *current* group assignment.

We calculated a *Need to Belong* score for each participant by reverse-coding the reverse-worded items and calculating an average of all the items on the NTB scale. Next we calculated the Pearson correlation between the EF index and the NTB index, r (53) = 0.29, p = .035 (see Fig. 6). In other words, there was a moderate relationship between individual's social identity goals and their change in implicit intergroup evaluations.

7.3. Discussion

In Experiment 3, we examined reversal of implicit evaluation within the same individual, in addition to the formation of implicit evaluation. We arbitrarily switched participants to what used to be their "out-group" halfway through the experiment, by simply telling them they were accidentally assigned to the wrong group. We replicated findings

Table 5
Mean response times (RTs) and standard deviations (Exp 3 after group switch).

	Target valence					
	Positive			Negative		
	M (SD)	95% CI		M (SD)	95% CI	
		Lower	Upper		Lower	Upper
Original in-group	685.56 (139.13)	647.21	723.91	669.34 (124.81)	634.94	703.75
Original out-group	689.36 (152.34)	647.37	731.35	691.29 (131.22)	655.12	727.46

Note: Experiment 3 Means and Standard Deviations of raw reaction times (RTs) for all four cells: 2 (in-group vs. out-group) x 2 (target valence: positive vs. negative) after group switch (N = 53). M = Means, SD = Standard Deviation, CI = Confidence Interval.

from Experiment 1 and Experiment 2 showing that participants developed implicit preferences towards their minimal in-group quickly. Moreover, also consistent with patterns from Experiment 1 and the competitive context in Experiment 2, the implicit intergroup preference prior to group switch appeared to be driven by an implicit out-group derogation. More importantly, we found that implicit intergroup preference could be quickly shifted as soon as the group assignments changed. After switching groups, participants reversed their implicit preferences, favoring the original out-group (or the new in-group). Participants in general responded more quickly to target images after switching groups, which is likely a result of practice and familiarity with the task in general (e.g., also see Hu, Rosenfeld, & Bodenhausen, 2012 for similar patterns).

Furthermore, we believed that participants' evaluative flexibility was not uniform. We examined the extent to which people readily shifted their implicit group preferences was associated with their individual belonging needs. As a result, participants' belonging needs positively predicted their evaluative flexibility – in other words, the extent to which they reversed their implicit group preferences after the group switch manipulation. These findings are interesting in that they not only speak to the fact that our implicit evaluations *can* be quickly shifted, but also point to a future direction of studying *who* tend to shift implicit intergroup preferences to a greater extent.

8. General discussion

In this current research, we examined whether implicit evaluations can form *and* change quickly through the influence of explicit goals and motives—specifically social identity goals—rather than only do so slowly through repeated associations or exposures. We also found that whether these newly formed implicit evaluations were sensitive to subtle changes in the intergroup context and associated with social goals. Together, these experiments provide evidence that social identities tune implicit evaluations in a flexible fashion.

Across three experiments, we replicated prior research showing that people quickly developed implicit preferences for their in-group, relative to their out-group, even though the groups were created and assigned on rather arbitrary basis. This pattern of results replicated and extended prior work on minimal social identities (e.g., Ashburn-Nardo et al., 2001; Van Bavel & Cunningham, 2009). Moreover, these implicit intergroup preferences appeared to be driven by implicit out-group derogation. This basic pattern was replicated in each experiment.

We then examined whether this quick formation of implicit evaluation was sensitive to the current intergroup context. When the two groups were described as competitive, we replicated the pattern of implicit intergroup preference from Experiment 1. However, when the two groups were described as cooperative, this pattern of implicit intergroup preference was eliminated. Importantly, participants only became aware of this intergroup context immediately before we assessed their implicit evaluations using our response time measure, making it impossible for participants to repeat or rehearse this information. This dynamic aspect of intergroup evaluation has not been observed in previously research, and offers insights into the process through which implicit evaluations develop.

Finally, we tested whether newly formed implicit evaluations would change to reflect sudden shifts in group identity. To test this possibility, we conducted an experiment where people were forced to switch teams—not unlike professional athletes being traded from one team to another or employees changing jobs. Following the group switch, implicit intergroup preferences were in the opposite direction as people favored their new in-group (even if they were out-group members mere moments earlier). This reversal of implicit group evaluations did not occur uniformly for everybody, but was predicted by individuals' own need to belong (see also Van Bavel et al., 2012). In other words, people who want to connect with others are the most flexible in their intergroup preferences, aligning their evaluations to fit in with their new group.

Prior to this research, there had been mixed findings regarding whether implicit evaluations could be *shaped* quickly by explicit goals once they form, or whether they are “easier done than undone” (e.g., Cone & Ferguson, 2015; Gregg et al., 2006, p. 1; Kurdi & Banaji, 2017). We found evidence of the flexibility of implicit group preferences. Specifically, implicit group preferences were observed immediately after group assignment and changed swiftly when the intergroup context become cooperative, or when the individual was suddenly assigned to the out-group. These findings provided clear evidence that newly formed implicit evaluations can and do change rapidly, even reversing, to reflect the current intergroup contexts and contingencies. At least when social identity is at stake, it seems that implicit evaluations can be easily done and undone.

Compared with previous research showing that existing implicit attitudes could *not* be changed quickly through explicit goals and motives, our current work differed in a few important ways. First, we examined implicit (group) evaluations that were formed on the spot using our minimal group paradigm (in this respect similar to Gregg et al., 2006) while some previous research focused on implicit attitudes with strong historical bases (e.g., implicit racial bias; e.g., Lai et al., 2014, 2016). While the latter forms of implicit attitudes certainly have practical significance and are important to study, they are a special form of implicit attitudes that are limited in terms of testing implicit attitudes per se. Our prior work suggests that even these longstanding group identities can be shifted when people develop superordinate identities (Van Bavel & Cunningham, 2009). However, it seems critical that social identity is leveraged to shape implicit preferences. When attempts to change implicit attitudes are up against systemic barriers that reinforce these identities, it may be difficult to sustain any changes in implicit bias.

Second, our manipulation relied on group assignment and identification, as compared to simply learning about novel groups (as in Gregg et al., 2006). Along similar lines, we speculate that the methods or interventions used in some prior research to change, reduce, or reverse implicit attitudes may not have induced sufficient and genuine goals among participants (e.g., using simple experiment instructions). It was therefore difficult to directly assess the extent to which each participant actually possessed the need/motive. In our research, we directly measured a motive that is fundamental to social preferences (i.e., group affiliative motivation). To our knowledge, this was the first research in this area that assessed relevant individual differences to predict ease of evaluative flexibility. We studied implicit evaluations in an intergroup context to take advantage of the strong fundamental motivation of group affiliation and identification.

Compared to previous research showing that newly formed implicit evaluations *can* change quickly, our findings also make unique contributions. Although different in their specific details, many previous studies on this topic have used impression formation type tasks (Brannon & Gawronski, 2017; Cone & Ferguson, 2015; Van Dessel et al., 2017). For instance, these researchers examined whether one-shot information that is contrary to previous learning could reverse previously formed implicit associations about a person or groups of people. Our research used a minimal group paradigm and assessed implicit group bias. It seems possible that this might allow for greater generalization since group preferences extend beyond individuals and can impact judgments and behaviors towards brand new individuals. Moreover, the minimal group approach captures a psychological process that is at the root of many intergroup preferences (see Dunham, 2018).

Our work on social identity and implicit evaluation also has interpretations for work on attitude change. Two recent, large-scale preregistered replication projects involving labs from across the country examined the impact of different interventions on implicit bias. For instance, White participants were assigned to a group with all Black ingroup members and all White out-group members (Lai et al., 2014, 2016). This group assignment was sufficient to generate positive

implicit evaluations towards in-group members—eliminating implicit racial biases. However, following up with these same participants 24–48 h later revealed that their implicit racial biases had returned, suggesting that once their group identity was no longer salient, the effects of group membership were easily undone (Lai et al., 2016). Thus, implicit attitudes developed and changed to follow social identity cues—but also to the removal of such cues—within the span of hours. Our work suggests that understanding the fluid nature of social identity will be critical in changing implicit bias.

9. Conclusion

The current research provides evidence of the flexibility of implicit evaluations—especially in an intergroup context. This work adds to the theoretical debate on implicit attitude formation and change, suggesting that implicit evaluations may be easier to change that previously believed. Yet this research may also help explain why implicit prejudice is so resistant to change. While our findings suggest that group identification and belonging motives can induce and reverse implicit evaluations, they may also sustain implicit prejudice in contexts where identities are stable and salient. Thus, factors like segregation, media narratives, and hierarchies that reinforce existing identities can undercut attempts to reduce implicit bias. This may help explain why intervention strategies and unconscious bias training aimed at changing implicit associations are often doomed to failure without an attempt to alter these structural and systemic factors. On the more promising side, this research may offer new directions for devising interventions to reduce or eliminate implicit prejudice, by targeting aspects of the local intergroup context.

Author contribution

YJX (yxiao2@uw.edu) and JVB (jay.vanbavel@nyu.edu) designed all experiments; YJX programmed and analyzed all experiments with input from JVB; YJX and JVB wrote the manuscript.

Acknowledgements

The authors would like to thank several research assistants (Annie Tak, Ming Yang, Julia Schaus, Amrita Balgobind, Francis Hwang) for data collection, and members of the Social Perception and Evaluation Lab at NYU for feedback on the manuscript. This research was partially funded by a National Science Foundation Grant to JVB (#1349089).

Appendix A. Collective identification scale

Please indicate the extent to which you agree or disagree with following the statements, by circling a number next to each statement.

	Strongly disagree			Neutral			Strongly agree
Being a member of the RED/BLUE team is important to my identity.	−3	−2	−1	0	1	2	3
I do not value being a member of the RED/BLUE team.	−3	−2	−1	0	1	2	3
I am proud to be a member of the RED/BLUE team.	−3	−2	−1	0	1	2	3

Appendix B. Need to Belong (NTB) scale

Please indicate the extent to which you agree or disagree with following the statements, by circling a number next to each statement.

	Strongly disagree			Neutral			Strongly agree
If other people don't accept me, I don't let it bother me.	−3	−2	−1	0	1	2	3
I try not to do things that will make other people avoid or reject me.	−3	−2	−1	0	1	2	3
I seldom worry about whether other people care about me.	−3	−2	−1	0	1	2	3
I need to feel that there are people I can turn to in times of need.	−3	−2	−1	0	1	2	3
I want other people to accept me.	−3	−2	−1	0	1	2	3
I do not like being alone.	−3	−2	−1	0	1	2	3
Being apart from my friends for a long period of time does not bother me.	−3	−2	−1	0	1	2	3
I have a strong need to belong.	−3	−2	−1	0	1	2	3
It bothers me a great deal when I am not included in other people's plans.	−3	−2	−1	0	1	2	3
My feelings are easily hurt when I feel that others do not accept me.	−3	−2	−1	0	1	2	3

References

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.). *A handbook of social psychology* (pp. 798–844). Worcester, MA: Clark University Press.

Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology, 81*(5), 789–799.

Baumeister, R. F., & Heatherton, T. F. (1996). Self-regulation failure: An overview. *Psychological Inquiry, 7*(1), 1–15.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529.

Bernstein, M. J., Young, S. G., Brown, C. M., Sacco, D. F., & Claypool, H. M. (2008). Adaptive responses to social exclusion: Social rejection improves detection of real and fake smiles. *Psychological Science, 19*(10), 981–983.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*(3), 242–261.

Bogardus, E. S. (1925). Measuring social distance. *Journal of Applied Sociology, 9*(2), 299–308.

Brannon, S. M., & Gawronski, B. (2017). A second chance for first impressions? Exploring the context-(in)dependent updating of implicit evaluations. *Social Psychological and Personality Science, 8*(3), 275–283.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86*(2), 307–324.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Scrull, & R. S. Wyer (Eds.). *Advances in social cognition* (pp. 1–36). Hillsdale, NJ: Erlbaum.

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin, 17*(5), 475–482.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues, 55*(3), 429–444.

Brosch, T., & Van Bavel, J. J. (2012). The flexibility of emotional attention: Accessible social identities guide rapid attentional orienting. *Cognition, 125*(2), 309–316.

Brown, J. D., Collins, R. L., & Schmidt, G. W. (1988). Self-esteem and direct versus indirect forms of self-enhancement. *Journal of Personality and Social Psychology, 55*(3), 445–453.

Chabris, C., & Brown, M. (2018, April 25). Starbucks's troubles can be a test for anti-bias training: Does it work? *Wall street journal*. Retrieved from https://www.wsj.com/articles/starbuckss-troubles-can-be-a-test-for-anti-bias-training-does-it-work-1524669057.

Cikara, M., & Van Bavel, J. J. (2014). The neuroscience of intergroup relations an integrative review. *Perspectives on Psychological Science, 9*(3), 245–274.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). HIllsdale, NJ: Erlbaum.
Cone, J., & Ferguson, M. J. (2015). He did what? The role of disgnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology, 108*(1), 37–57.
Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition, 25*(5), 736–760.
Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81*(5), 800–814.
Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18.
Devine, P. G., Forscher, P. S., Austin, A., & Cox, W. T. L. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology, 48*(6), 1267–1278.
Dotsch, R., Wigboldus, D., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of Personality and Social Psychology, 100*, 999–1014.
Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*(1), 62–68.
Dunham, Y. (2018). Mere membership. *Trends in Cognitive Sciences, 22*(9), 780–793.
Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* Orlando, FL: Harcourt Brace Jovanovich.
Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.
Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobstrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*(6), 1013–1027.
Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*(1), 297–327.
Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*(2), 229–238.
Ferguson, M. J., & Bargh, J. A. (2004). Liking is for doing: The effects of goal pursuit on automatic evaluation. *Journal of Personality and Social Psychology, 87*(5), 557–572.
Galdi, S., Arcuri, L., & Gawronski, B. (2008). Automatic mental associations predict future choices of undecided decision-makers. *Science, 321*(5892), 1100–1102.
Gardner, W. L., Pickett, C. L., & Brewer, M. B. (2000). Social exclusion and selective memory: How the need to belong influences memory for social events. *Personality and Social Psychology Bulletin, 26*(4), 486–496.
Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE Model. *Social Cognition, 25*(5), 687–717.
Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology, 40*, 535–542.
Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4.
Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480.
Gregg, A., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preference. *Journal of Personality and Social Psychology, 90*(1), 1–20.
Hastorf, A. H., & Cantril, H. (1954). They saw a game: A case study. *Journal of Abnormal and Social Psychology, 49*(1), 129–134.
Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup Bias. *Annual Review of Psychology, 53*, 575–604.
Honkanen, P., Verplnken, B., & Olsen, S. O. (2006). Ethical values and motives driving organic food choice. *Journal of Consumer Behaviour, 5*, 420–430.
Hu, X., Antony, J. W., Creery, J. D., Vargas, I. M., Bodenhausen, G. V., & Paller, K. A. (2015). Unlearning implicit social biases during sleep. *Science, 348*(6238), 1013–1015.
Hu, X., Rosenfeld, J. P., & Bodenhausen, G. V. (2012). Combating automatic autobipgraphical associations: The effect of instruction and training in strategically concealing information in the autobiographical implicit association test. *Psychological Science, 23*(10), 1079–1085. https://doi.org/10.1177/0956797612443834.
Joy-Gaba, J. A., & Nosek, B. A. (2010). The surprisingly limited malleability of implicit racial evaluations. *Social Psychology, 41*(3), 137–146.
Jussim, L. (2017). Mandatory Implicit Bias Training Is a Bad Idea. Retrieved from https://www.psychologytoday.com/us/blog/rabble-rouser/201712/mandatory-implicit-bias-training-is-bad-idea
Kurdi, B., & Banaji, M. R. (2017). Repeated evaluative pairings and evaluative statements: How effectively do they shift implicit attitude? *Journal of Experimental Psychology: General, 146*(2), 194–213.
Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., ... Banaji, M. R. (2018). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *The American Psychologist.* https://doi.org/10.31234/osf.io/582gh.
Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. L., Joy-Gaba, J. A., ... Nosek, B. A. (2014). Reducing implicit racial preference: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General, 143*, 1765–1785.
Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., ... Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General, 145*, 1001–1016.
Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical report A-8*.
Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24*(8), 1377–1388.
Leary, M. R., Kelly, K. M., Cottrell, C. A., & Schreindorfer, L. S. (2013). Construct validity of the Need to Belong scale: Mapping the nomological network. *Journal of Personality Assessment, 95*(6), 610–624.
Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81*(5), 842–855.
Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51*, 93–120.
Maio, G. R., & Haddock, G. (2015). *The psychology of attitudes and attitude change* (2nd ed.). Thousand Oaks, CA: Sage.
Maison, D., Greenwald, A. G., & Bruin, R. H. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology, 14*(4), 405–415.
Maner, J. K., DeWall, C. N., & Baumeister, R. F. (2007). Does social exclusion motivate interpersonal reconnection? Resolving the "Porcupine Problem". *Journal of Personality and Social Psychology, 92*(1), 42–55.
Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluation. *Journal of Personality and Social Psychology, 108*(6), 823–849.
Nosek, B. A., & Banaji, M. R. (2001). The Go/No-Go Association Task. *Social Cognition, 19*(6), 625–664.
Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*(5), 413.
OSC. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.
Otten, S., & Moskowitz, G. B. (2000). Evidence for implicit evaluative in-group bias: Affect-biased spontaneous trait inference in a minimal group paradigm. *Journal of Experimental Social Psychology, 36*(1), 77–89.
Penner, L. A., Dovidio, J. F., West, T. V., Gaertner, S. L., Albrecht, T. L., Dailey, R. K., & Markova, T. (2010). Aversive racism and medical interactions with Black patients: A field study. *Journal of Experimental Social Psychology, 46*(2), 436–440.
Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classical and contemporary approaches.* Dubuque, IA: Brown.
Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Vol. Ed.), *Advances in experimental social psychology. Vol. 19. Advances in experimental social psychology* (pp. 123–205). New York: Academic Press.
Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology, 48*, 609–647.
Pickett, C. L., Gardner, W. L., & Knowles, M. L. (2004). Getting a cue: The need to belong and enhanced sensitivity to social cues. *Personality and Social Psychology Bulletin, 30*(9), 1095–1107.
Postmes, T., & Jetten, J. (Eds.). (2006). *Individuality and the group: Advances in social identity*. London: Sage.
Ratner, K. G., & Amodio, D. M. (2013). Seeing "us vs. them": Minimal group effects on the neural encoding of faces. *Journal of Experimental Social Psychology, 49*, 298–301.
Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91*(6), 995–1008.
Scroggins, W. A., Mackie, D. M., Allen, T. J., & Sherman, J. W. (2016). Reducing prejudice with labels: Shared group memberships attenuate implicit bias and expand implicit group boundaries. *Personality and Social Psychology Bulletin, 42*(2), 219–229.
Sinclair, S., Lowery, B. S., Hardin, C. D., & Colangelo, A. (2005). Social tuning of automatic racial attitudes: The role of affiliative motivation. *Journal of Personality and Social Psychology, 89*(4), 583.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3.
Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*(2), 108.
Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220–247.
Tajfel, H. (1982). *Social identity and intergroup relations.* Cambridge, UK: Cambridge University Press.
Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology, 1*, 149–178.
Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33*(4), 529–554.
Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory.* Cambridge, MA: Blackwell.
Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin, 20*, 454–463.
Van Bavel, J. J., & Cunningham, W. A. (2009). Self-categorization with a novel mixed-race group moderates automatic social and racial biases. *Personality and Social Psychology Bulletin, 35*(3), 321–335.
Van Bavel, J. J., & Cunningham, W. A. (2012). A social identity approach to person memory group membership, collective identification, and social role shape attention and memory. *Personality and Social Psychology Bulletin, 38*(12), 1566–1578.
Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of in-group bias. *Psychological Science, 19*(11), 1131.
Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2011). Modulation of the fusiform face area following minimal exposure to motivationally relevant faces: Evidence of

in-group enhancement (not out-group disregard). *Journal of Cognitive Neuroscience, 23*(11), 3343–3354.
Van Bavel, J. J., Swencionis, J. K., O'Connor, R. C., & Cunningham, W. A. (2012). Motivated social memory: Belonging needs moderate the own-group bias in face recognition. *Journal of Experimental Social Psychology, 48*(3), 707–713.
Van Dessel, P., Mertens, G., Smith, C. T., & De Houwer, J. (2017). The mere exposure instruction effect. *Experimental Psychology, 64*(5), 299–314. https://doi.org/10.1027/1618-3169/a000376.
Van Dessel, P., Ye, Y., & De Houwer, J. (2018). Changing deep-rooted implicit evaluation in the blink of an eye. *Social Psychological and Personality Science, 9*, 1–8.
Weisbuch, M., & Ambady, N. (2008). Affective divergence: Automatic responses to others' emotions depend on group membership. *Journal of Personality and Social Psychology, 95*(5), 1063–1079.
Williams, K. D., & Sommer, K. L. (1997). Social ostracism by coworkers: Does rejection lead to loading or compensation? *Journal of Personality and Social Psychology, 23*(7), 693–706.
Xiao, Y. J., Coppin, G., & Van Bavel, J. J. (2016). Perceiving the world through group-colored glasses: A perceptual model of intergroup relations. *Psychological Inquiry, 27*(4), 255–274.
Young, A. I., Ratner, K. G., & Fazio, R. H. (2014). Political attitudes bias mental representation of a presidential candidate's face. *Psychological Science, 25*, 503–510.