The Impact of Loyalty and Equality on Implicit Ingroup Favoritism

Cristina Zogmaister, Luciano Arcuri, Luigi Castelli and Eliot R. Smith

Group Processes Intergroup Relations 2008 11: 493
DOI: 10.1177/1368430208095402

Group Processes & Intergroup Relations
2008 Vol 11(4) 493–512

GPIR

The Impact of Loyalty and Equality on Implicit Ingroup Favoritism

Cristina Zogmaister
University of Milan-Bicocca

Luciano Arcuri and Luigi Castelli
University of Padova, Italy

Eliot R. Smith
Indiana University

Extending recent investigations into the malleability of implicit ingroup favoritism, three experiments examined the role of indirect activation of equality and loyalty. Results showed that priming equality decreased implicit favoritism, measured through the Implicit Association Test and Go/No-Go Association Task, whereas priming loyalty enhanced it; spontaneous behavior (seating distance) was similarly influenced. A boundary condition was observed, namely change of intergroup setting: the effects of priming equality and loyalty ceased when these were primed after an irrelevant ingroup identity was made salient. In general, implicit favoritism can be reduced or increased after the activation of equality and loyalty respectively, and this underlines the importance of tackling discrimination by both lessening its expression, and removing factors that exacerbate it.

KEYWORDS equality, implicit ingroup favoritism, intergroup behavior, loyalty, prejudice malleability

In recent years, a lot of effort has been devoted to the investigation of contextual moderation of automatic intergroup bias (Amodio & Devine, 2006; Blair, 2002). The bulk of research has focused on implicit stereotyping, which may be defined as the automatic activation and application of semantic knowledge related to a certain social category (reviewed by Blair, 2002; see Castelli, Macrae, Zogmaister, & Arcuri, 2004; Dambrun & Guimond, 2004; Moskowitz, Li, & Kirk, 2004; Sassenberg & Moskowitz, 2005 for more recent studies). This set of studies demonstrates that various kinds of manipulations influence automatic stereotyping, including: (i) mental imagery (Blair, Ma, & Lenton, 2001; but see Gawronski & Bodenhausen, 2005); (ii) training to think in a counter-stereotypic way (Gawronski, Deutsch, Mbirkou, Seibt, & Strack, 2008; Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000); (iii) specific characteristics of the target stimuli or the way they are presented (e.g. Castelli et al., 2004; Macrae, Bodenhausen, & Milne, 1995; Macrae, Mitchell, & Pendry, 2002); (iv) processing

Author's note
Address correspondence to Cristina Zogmaister, who is now at the Department of Psychology, University of Milan-Bicocca, Piazza dell' Ateneo Nuovo, 1, I-20126 Milan, Italy [email: cristina.zogmaister@unimib.it]

11:4; 493–512; DOI: 10.1177/1368430208095402

goals (Macrae, Bodenhausen, Milne, Thorn, & Castelli, 1997; see also Gollwitzer & Schaal, 1998); (v) creativity goals (Sassenberg & Moskowitz, 2005); (vi) learning that consensus about one's stereotypes is low (Sechrist & Stangor, 2001); (vii) motivation to form a certain impression of the target (Sinclair & Kunda, 1999); and (viii) threats to one's self-image (Spencer, Fein, Wolfe, Fong, & Dunn, 1998).

Studies investigating contextual features able to moderate implicit ingroup favoritism, the evaluative component of automatic reactions toward outgroups, are scarcer. In particular, research has found that the expression of implicit ingroup favoritism and prejudice is influenced by the presentation of exemplars of the stigmatized group that are liked (Dasgupta & Greenwald, 2001) or are associated with specific settings (Barden, Maddux, Petty, & Brewer, 2004), by the presence of persons (allegedly) holding egalitarian views and consequent social tuning demands (Richeson & Ambady, 2003; Sinclair, Lowery, Hardin, & Colangelo, 2005), and by the request to give profound consideration to multi-culturalist values (Richeson & Nussbaum, 2004).

It is important to underline that we cannot simply draw a parallel between those factors moderating implicit stereotyping and those moderating implicit ingroup favoritism. Stereotyping and ingroup favoritism are conceptualized as qualitatively different (Park & Judd, 2005). Research has shown that these two forms of bias often do not correlate, and even more importantly, implicit ingroup favoritism and implicit stereotyping appear to be related to different kinds of behaviors (see Amodio & Devine, 2006, for a thorough discussion).

The present study

The basic goal of the present study is to enhance knowledge of the conditions that influence implicit favoritism, focusing on the contextual activation of general goals of equality and loyalty. Previous studies have indeed shown the importance of egalitarian goals for the control of implicit stereotyping (Moskowitz, Gollwitzer, Wasel, & Schaal, 1999; Moskowitz, Salomon, & Taylor, 2000). In two related sets of studies, Moskowitz and his colleagues demonstrated that chronically egalitarian participants automatically activated egalitarian goals upon encountering a member of a stigmatized group, and inhibited stereotype-related information.

The present study expands this line of research in several ways. First, we focused on *contextual*, as opposite to chronic, goals, and on implicit *ingroup favoritism*, instead of stereotyping. In addition, the influence of goals on spontaneous intergroup *behavior* was investigated and we extend the scope of research by including both the social goal of equality, which was hypothesized to decrease intergroup bias, and the goal of loyalty which, by contrast, was hypothesized to *increase* bias.

Contextual enhancement of implicit favoritism

As mentioned, an important question we want to address here is whether there is the risk that specific mental contents might further increase implicit ingroup bias over its baseline. To our knowledge, to date, no study has specifically investigated this question. In some of the studies on the contextual modulation of evaluative bias, an experimental condition was present in which ingroup bias was attenuated, as compared with a neutral control condition (Richeson & Ambady, 2003; Sinclair et al., 2005). In other studies, an experimental condition expected to decrease ingroup bias was compared with another condition expected to enhance it (Barden et al., 2004; Dasgupta & Greenwald, 2001; Richeson & Nussbaum, 2004). To date, no study has contrasted a favoritism-enhancing condition with a control condition. Consequently, there is evidence on the decrease of implicit favoritism by specific contextual manipulations, but it is not clear whether there are conditions that risk further intensifying implicit favoritism.

It is both theoretically and practically important to understand what factors enhance ingroup bias. From the theoretical point of view it is conceivable that people express a high level of evaluative and behavioral discrimination against outgroups, in order to increase the positivity of their social identity (Tajfel, 1981; Tajfel & Turner, 1986). But do conditions exist in which the level of evaluative bias may even worsen? The present

research will seek to answer this question. From the practical point of view, knowing which factors may increase intergroup bias could alert decision makers and guide prevention policies both at the broader level, as in the case of religious and interethnic conflicts, and at the narrower level, as in the case of sports competitions between long-standing rival teams.

For the enhancement of favoritism, we focused on the goal of loyalty. Research on the human value system shows that loyalty is an important value, largely shared among individuals (Braithwaite & Law, 1985; Crace & Brown, 1996; Rokeach, 1967, 1973; see also Scott, 1965). The importance of loyalty for the existence of social groups is stressed by the work of van Vugt and Hart (2004), who hypothesize and bring empirical support to the notion that loyalty may function as a 'social glue' that holds groups together, and contributes to group stability and integrity.

As noted by van Vugt and Hart (2004, p. 585), 'loyalty is a complex, multifaceted construct, consisting of emotive, cognitive, as well as behavioral elements'. Loyal behaviors are those involving personal sacrifice for the interests of the group (Levine & Moreland, 2002; van Vugt & Hart, 2004). Emotional components may be a precursor of behavioral loyalty (van Vugt & Hart, 2004). They may also emerge as responses to loyal and disloyal behaviors performed by the individual (Zdaniuk & Levine, 2001) and by other ingroup members (Abrams, Marques, Bown, & Dougill, 2002; Abrams, Marques, Bown, & Henson, 2000; Castelli, Tomelleri, & Zogmaister, 2008; van Vugt & Chang, 2006).

Loyalty has a positive connotation, but studies on the consequences of national identification suggest that it may also have negative implications for intergroup relations. Indeed, research conducted by Li and Brewer (2004) indicates that, in certain circumstances, being a highly identified and loyal member of the ingroup is related to nationalistic attitudes, characterized by beliefs in the superiority of the ingroup (see also Blank & Schmidt, 2003). Therefore, we hypothesized that activating the value of loyalty could cause an increase in intergroup bias.

A previous minimal group study by Hertel and Kerr (2001) demonstrated that priming the concept of loyalty indeed increased ingroup favoritism and identification, as compared with a condition in which equality was primed. Because no control condition was present in Hertel and Kerr's study, it is impossible to assess the relative contribution of equality and loyalty priming. In the present set of studies, a control condition will help to disentangle the effect of loyalty from the effect of equality.

Modulation of intergroup behavior

Understanding what contextual factors modulate automatic intergroup evaluations is particularly important for the ultimate influence of these attitudes on intergroup behavior. Measures of implicit evaluative bias correlate with spontaneous behaviors toward members of racial outgroups, particularly with those behaviors that fall outside conscious control and those that people usually do not view as an indication of their attitude and thus do not try to control (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; see Strack & Deutsch, 2004). Knowing which factors decrease or increase ingroup bias is valuable because parallel effects of these same factors are expected on overt intergroup behavior. However, a more stringent test of the importance of the activation of goals of equality or loyalty for intergroup relations will be the direct investigation of the influence of these factors on real intergroup behavior. It is important to note that, even though actors are often unaware of the subtle bias represented in their spontaneous behavior, targets may be well aware of it. To the extent that these subtle manifestations of racial bias are detected, they negatively affect intergroup relations (Dovidio, Kawakami, & Gaertner, 2002). Therefore, it is important to identify contextual factors that influence spontaneous intergroup behaviors.

In particular, we considered seating distance, which is a spontaneous aspect of nonverbal intergroup behavior that has proven useful in previous research (see Amodio & Devine, 2006; Hendricks & Bootzin, 1976; Macrae, Bodenhausen, Milne, & Jetten, 1994). Interpersonal distance is known to be related to interpersonal relationships and attitudes; it is minimal in intimate interactions, intermediate in friendships, and higher in cases of superficial knowledge and work relations

(Argyle & Dean, 1965; Byrne, 1961). The interpersonal distance in an interaction is an aspect of interpersonal behavior that often goes uncontrolled, and hence intergroup bias may surface in seating distance even for well-intentioned people. In the case of intergroup relations, distance is associated with the degree of discomfort during interaction (Hendricks & Bootzin, 1976), and can be considered as an indication of the level of intimacy in a social interaction and of the attitude toward the interaction partner. It has been shown that people often prefer a larger distance between themselves and those belonging to stigmatized groups (Barrios, Corbitt, Estes, & Topping, 1976).

Overview We report three experiments designed to investigate the effect of the activation of the two intergroup goals of loyalty and equality on implicit favoritism and on spontaneous intergroup behavior. The following questions are addressed: Is the activation of the goal to be equal or to be loyal sufficient to modulate implicit favoritism (Experiment 1)? Is the activation of an overarching category a boundary condition of this effect (Experiment 2)? Does the influence extend to spontaneous intergroup behavior (Experiment 3)? In the experiments, intergroup goals were primed through the scrambled-sentence task (Srull & Wyer, 1979), and a control condition was included in each study to assess the directionality of the effects.

To assess the generality of the hypothesized effects we focused on three different ingroup/outgroup differentiations: national, local and religious. Two different measures of implicit intergroup evaluation were used: the Implicit Association Test (IAT; Greenwald, McGhee, & Schwarz, 1998), and the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001); to investigate the effects of priming on intergroup behavior, seating distance was considered. A meta-analysis will be presented to statistically summarize results.

Experiment 1

In Experiment 1, we tested the hypothesis that priming goals of equality and loyalty would influence spontaneous expressions of implicit favoritism. After completing a scrambled-sentence task in one of three different conditions (equality, loyalty, and control), participants completed an IAT for the assessment of the automatic preference of their national ingroup (Italians), as compared with an outgroup (Germans). We predicted that, in line with previous research, an implicit preference for the ingroup over the outgroup would emerge. We further hypothesized that intergroup bias would be minimal for participants assigned to the equality condition, maximal for participants assigned to the loyalty condition, whereas the level of intergroup bias of participants in the control condition would fall in between. Third, we hypothesized that the effect of indirect priming would emerge even if participants were unaware that the priming task was aimed at activating a specific goal.

Method

Participants Ninety-six first-year Italian psychology students (90 females, 6 males) took part in the experiment in exchange for partial course credit. One third of participants were randomly assigned to the loyalty condition, one third to the equality condition, and one third to the control condition.

Materials

Scrambled-sentence task The priming procedure involved a 32-item scrambled-sentence task and was presented as a test of language processing. Three versions of the task were constructed (equality, loyalty, control version). In the loyalty version, 25 sentences made reference to the idea of supporting and helping members of the ingroup (e.g. 'L. helps his teammates', 'One's family members usually take priority'), and the remaining seven were neutral (referring neither to equality nor loyalty). Twenty-five of the 32 sentences in the equality version made reference to equality (e.g. 'All are equal before the law', 'C. fights for equality between the people'), and the remaining 7 were neutral. In the control version all 32 sentences were neutral (e.g. 'That politician is involved in the issue of pollution'). Seven sentences were the same as those used as

filler in the other two conditions, and 25 were different neutral sentences. Each trial consisted of five fragments, and participants were instructed to construct a grammatically correct four-fragment sentence as quickly as possible, leaving one of the fragments unemployed. The first sentence was neutral, and the subsequent sentences were presented in a fixed random order. The sentences were pretested with a sample of 10 respondents, who were presented with the list of all 82 sentences comprised in the scrambled-sentence task, in a fixed random order. They were asked to evaluate the meaning of each sentence on a 5-point scale ranging from 1 (*Extremely related to intergroup equality*), to 5 (*Extremely related to loyalty to one's own group*). The midpoint was labeled as neutral value. This pretest showed that the mean evaluations of the 7 filler and of the 25 neutral sentences were not different from the midpoint of the scale, all t values < 1. The 25 sentences related to loyalty were judged as significantly different from the midpoint of the scale and polarized toward the 'ingroup loyalty' pole, $M = 4.11$, $t(9) = 14.33$, $p < .001$. The 25 sentences related to equality were judged as significantly polarized toward the 'intergroup equality' pole of the scale, $M = 1.77$, $t(9) = 11.10$, $p < .001$.

Implicit Association Test To increase the indirect character of the measure, Italian and German cities were presented as stimuli representing the ingroup and the outgroup, instead of members of the ingroup and the outgroup (Kühnen et al., 2001). Stimuli for the IAT were names of eight well-known Italian and eight well-known German cities (e.g. Florence, Milan, Cologne, Berlin), together with six positive and six negative words from the original set of Greenwald et al. (1998) (e.g. joy, love, awful, evil). The IAT is a computer-administered task, in which participants are requested to categorize stimuli belonging to four different categories, by pressing two different keys of the keyboard. Participants are administered a sequence of five blocks of trials, three of which are learning blocks and the other two are critical. The implicit attitude index is based on the difference in participants' performance in these critical blocks. In one critical block, participants had to classify Italian cities and positive words by pressing one response key, and German cities and negative words by pressing the other key. In the other critical block, Italian cities and negative words shared one response key, and German cities and positive words shared the other response. The order of the two critical blocks was counterbalanced between participants. Participants were presented with 12 trials in the attribute-categorization learning block, 16 trials in the concept-categorization learning blocks, and 48 trials in the critical blocks.

Procedure The experiment was presented as a study of the cognitive mechanisms involved in language processing and categorization. Participants took part in the study in individual sessions of about 15 minutes. They first completed the scrambled-sentence task, in one of three different versions, depending on the condition (loyalty, equality, control). Immediately afterwards, participants were seated in front of a computer and performed the IAT. Thereafter, participants in the loyalty and equality condition were administered a short questionnaire in order to investigate awareness of the study purposes. Three questions were asked: (i) 'What is the aim of this experiment, in your opinion?'; (ii) 'What is the relationship between the first task (the paper questionnaire) and the second task (the computer-administered categorization task)?'; (iii) 'Do you think it is possible that the first task influenced your responses in the second one? (If yes) what was the influence?'. Then, participants were thanked and fully debriefed.

Results and discussion

Awareness questionnaire Three participants showed awareness that the experiment was related to favoritism. The remaining participants answered in ways related to our cover story (i.e. when asked about the aims of the experiment, they responded that it was about cognitive processes of thinking) or they wrote that they did not know. Importantly, although the third question of the questionnaire made direct reference to whether there was a connection

between the priming task and the measure of implicit associations, which could have alerted participants to the relationship, only one participant (in the equality condition) gave an affirmative answer. Analyses were conducted both including and discarding data from these aware participants, but no difference emerged in the pattern of results. Hence, results from analyses including all participants will be presented. Because of this lack of awareness of the influence of the scrambled-sentence task on performance in the IAT, any observed differences in mean IAT scores can be attributed to an implicit consequence of goal activation (Bargh, 1992).

Implicit favoritism Mean performance measures for the two combined tasks of the IAT were computed following the procedure described by Greenwald et al. (1998). We computed scores so that positive values indicate a preference for Italian over German cities and negative values indicate a preference for German cities. The IAT index had an acceptable internal consistency, $\alpha = .68$. The mean value was 239.23, $SD = 152.33$, indicating a significant ingroup bias, $t(95) = 15.39$, $p < .001$, $d = 1.57$. A 3 (primed goal: Equality/Control/Loyalty) × 2 (order of blocks: Italian-Positive vs. German-Positive first) analysis of variance (ANOVA) on the IAT index revealed only a main effect of the primed goal, $F(2, 90) = 3.13$, $p < .05$, $\eta^2_p = .07$. Supporting our predictions, a significant linear trend indicated an increase of ingroup bias from the equality to the control and from the control to the loyalty condition, $F(1, 93) = 6.75$, $p = .011$. Mean values of the IAT index depending on the priming condition are reported in Table 1.[1]

The linear trend analysis supports our predictions of a decrease in favoritism in the equality condition, and its increase in the loyalty condition, as compared with the control condition. This effect of goal activation can be considered automatic in two ways: because a measure of implicit attitudes was used and because participants were unaware of the consequences of priming on their performance on this measure. These results provide a first indication of the influence of the activation of a *temporary* goal on implicit favoritism. Furthermore, it demonstrates that loyalty concerns may increase intergroup discrimination, providing initial evidence for the hypothesis that implicit ingroup bias can be enhanced by the context. Because of the potential importance of these

Table 1. Means (*SD*) of the dependent variables, depending on the activated goal

	Equality	Control	Loyalty
Experiment 1			
Implicit intergroup bias	196.9_{a}	232.50_{ab}	288.29_{b}
	(147.76)	(130.84)	(166.66)
Experiment 2			
Implicit intergroup bias			
relevant identity condition	$.36_{a}$	2.00_{b}	2.44_{b}
	(1.58)	(1.06)	(1.83)
irrelevant identity condition	2.24_{a}	1.57_{a}	1.80_{a}
	(1.06)	(1.72)	(1.76)
Experiment 3			
Implicit identification	$-.11_{a}$	$-.26_{a}$	$.10_{a}$
	(.76)	(.99)	(.89)
Implicit intergroup bias	-3.54_{a}	130.97_{b}	176.04_{b}
	(190.16)	(221.77)	(268.53)
Explicit intergroup bias	$.20_{a}$	$.30_{a}$	$-.20_{a}$
	(.89)	(.92)	(1.01)
Seating distance	120.75_{a}	129.25_{ab}	141.25_{b}
	(23.91)	(16.88)	(23.16)

Note: Means in the same row with no subscript in common are statistically different, with an alpha level of .05.

results, we judged it necessary to provide a replication. Furthermore, we felt it important to investigate whether boundary conditions to this effect exist.

Experiment 2

So far, the results of the Experiment 1 indicated that the degree of ingroup bias displayed by participants was influenced by the goal that the context made salient. These results are consistent with the concept of mental representations as context-sensitive states (Smith & Conrey, 2007; see also Conrey & Smith, 2007). This notion is based on a distributed connectionist model of the memory system, in which representations are conceived as patterns of activation in a network of interconnected nodes.

The basic idea behind the notion of representations as context-sensitive states is that the representation formed by the network in response to a certain stimulus is jointly determined by the stimulus itself, and by the most relevant contextual information. Internal states like goals and motives are among the most relevant features of the social context (Smith & Conrey, 2007), and therefore they should have an important influence on the way representations are reconstructed.

In the General Discussion we briefly address the consequences of considering representations as states for the interpretation of attitude malleability. Here we want to emphasize that not only the representation of concepts, but also the representations of goals and motives can be conceived as context sensitive. Because the differentiation between ingroup and outgroup is one of the most important aspects of the social context, the activation of loyalty and equality while a specific group membership is salient should result in the formation of a context-specific representation of these concepts. For example, if the context enhances the salience of one's own identity as Northern Italian, and the notion of loyalty is activated, the representation of loyalty to Northern Italians may be formed, rather than a generic representation of loyalty. Similarly, if in the same context the notion of equality is activated, the representation of equality between Northern Italians and the natural outgroup of Southern Italians may be formed.

If this is the case, the activated notion of loyalty (or equality) should influence spontaneous attitudes toward the relevant groups (in the example, Northern and Southern Italians), but should not impact on attitudes toward irrelevant groups (e.g. one's occupational group). Therefore, a change in intergroup context between the moment of goal activation and the moment of expression of intergroup bias may constitute a boundary condition for the effect of equality and loyalty priming on ingroup favoritism. This hypothesis was tested in Experiment 2.

For the sake of generalizability, in Experiment 2 we investigated implicit evaluative bias through a different measure, developed a different set of sentences for the activation of the goals, and considered attitudes toward regional instead of national groups. In particular, the distinction between Northern and Southern Italians was explored, which is particularly pronounced and holds a particular weight for Italian citizens, due to the long history of deeply defined, distinct regional cultures within Italy. In fact, many Italians identify primarily with their region of origin, and view those from other areas of Italy as part of the outgroup (see Sanchez, Zogmaister, & Arcuri, 2007).

In Experiment 2 (unlike Experiment 1, in which no identity was activated before the scrambled-sentence task), either the regional identity (i.e. Northern Italians) or the identity as a student was initially made salient by way of a questionnaire. Thereafter, the goal was activated (loyalty, equality, or no goal, depending on the condition) through a scrambled-sentence task. Finally, the implicit attitude toward Northern and Southern Italians was assessed through the GNAT. We hypothesized an implicit preference for the ingroup. We furthermore expected loyalty priming to increase ingroup favoritism, and equality priming to decrease it, as compared with the control condition, but only for those participants who had their regional identity made salient at the beginning of the experiment.

Method

Participants Participants were 88 Italian students (71 females, 17 males). They received partial course credit for their participation.

Materials

Scrambled-Sentence Task For this experiment we developed three new sets of sentences. Whereas in the sets used in Experiment 1 the whole meaning of the sentences made reference either to loyalty to the own group or to intergroup equality, in each of the new sentences a single word was embedded, which was strictly related either to equality (egalitarian, equality, equal, fair, fairness, honest, impartial, neutral, parity, unprejudiced) or to loyalty (dependable, reliable, cohesive, group-spirited, loyal, disposed, loyalty, support, trustworthy, buddy). These words were either synonyms of the concept to-be-activated, or related to the concept according to the dictionary definition. Three different versions of the task were created (equality, loyalty, and control), each constituted by 14 sentences. Four of the sentences, two at the beginning and two at the end of the task, were the same in all conditions, and were neutral. The remaining 10 sentences differed according to the experimental condition.

Go/No-Go Association Task As materials for the GNAT, a series of 10 Northern (e.g. Venice, Milan) and 10 Southern (e.g. Naples, Palermo) Italian cities of different size and prominence were chosen, together with a series of 10 positive (e.g. excellent, happy) and 10 negative (e.g. bad, horrible) words. The GNAT was computer-administered. It consisted of a series of eight blocks of trials. In each trial, a stimulus was presented on the monitor, and participants were requested to decide as fast and accurately as possible whether it belonged to a given category. When the stimulus was a member of the category, participants had to press the space bar within 660 ms from stimulus onset. When the stimulus did not belong to the category, participants were instructed to make no response. The first series of four blocks allowed participants to learn the response mode for each category of stimuli. The remaining four blocks were critical. In each critical block, participants were requested to react to members of two target categories. For example, in one critical block the target categories were 'Southern Italian', and 'positive'. Participants had to press the space bar every time either the name of a Southern Italian city or a positive word appeared on the monitor, whereas when a Northern Italian city or a negative word appeared, they should not press the space bar. The short time allowed for the response was expected to cause a high number of errors, particularly when the association was counter-attitudinal (e.g. for a Northern Italian respondent, when Northern Italian cities required the same response as negative words). Participants were presented with 12 trials in the learning blocks, and with 68 'true' trials in the critical blocks. These true trials, in which data were recorded and analyzed, were preceded by 10 'warm-up' trials; performance in these warm-up trials was not included in the analyses. An interval between the warm-up and the true trials allowed respondents to ask for possible clarifications. The order of tasks was randomized within the first series of four practice blocks, and within the second series of four critical blocks.

Procedure Participants were individually tested. Upon entering the laboratory, they were randomly assigned to one of two identity-activation conditions: relevant and irrelevant. For the relevant identity condition, at the beginning of the experiment participants were asked whether they were Northern or Southern Italians and subsequently received a sheet of paper containing the following instructions: 'Write down four characteristics that make you feel like a Northern Italian' (or Southern Italian, according to their regional origin). For the irrelevant identity condition, no mention was made of regional origin and participants received a sheet of paper containing the instructions to write down four characteristics that made them feel like students. After answering this task, participants completed the scrambled-sentence task, according to the condition they were assigned to, and then went through the GNAT. Finally, they were thanked for participation and fully

debriefed. During debriefing, participants were asked about their guesses regarding the hypotheses behind the study and the aim of the scrambled-sentence task. Participants in the irrelevant identity condition were asked at this point about their Northern or Southern membership. Subsequently, they were given a full explanation concerning the study and our main hypotheses. No participant showed awareness of our hypotheses or guessed the aims of the scrambled-sentence task. The whole procedure lasted about 20 min.

Results and discussion

Data from 14 participants were discarded because they were of Southern Italian origin,[2] and data from 8 Northern Italian participants were discarded because they failed to comply with instructions.[3] Hence, analyses were conducted on data from 66 participants. No participant expressed any suspicion during the debriefing interview about the possible influences of the initial questionnaire or the scrambled-sentence task on their reactions in the GNAT.

Following Nosek and Banaji (2001), signal detection analysis was conducted. In the first step, we separately computed the *d'* index of discriminability and the β criterion for the detection of cities and words in each of the four critical blocks. We subsequently investigated the stability of the β criterions for the detection of cities and valenced words. It emerged that the β criterion for the detection of valenced words was significantly influenced by the experimental condition, whereas β values for the detection of cities were stable across experimental conditions. This means that participants chose different criteria for recognizing that a valenced word belonged to the target category, depending on the normative condition, whereas the criterion remained stable when a city was the target. The index of discriminability *d'* can be influenced by variations in the criterion, and therefore the same detectability levels can be accompanied by different *d'* indexes when the criterion fluctuates (see Swets, 1986). For this reason, we performed our main analyses on the *d'* indexes for the detection of cities, in order to remove this source of systematic error. A very similar pattern of results, although characterized by lower effect sizes, emerged from parallel analysis conducted on *d'* indexes computed for the detection of all stimuli.[4]

In order to compute a differential index of intergroup bias, we summed the *d'* index of the ingroup cities/positive words and outgroup cities/negative words blocks, and (separately) the *d'* index of the outgroup cities/positive words and ingroup cities/negative words blocks. We then subtracted the second value from the first in order to obtain a differential index of automatic preference for the ingroup. Hence, higher values represent a stronger automatic preference for the ingroup. The GNAT index had a mean value of 1.74, $SD = 1.63$, indicating a significant ingroup bias, $t(65) = 8.68$, $p < .001$, $d = 1.07$.

This index of intergroup bias was submitted to a 2 (activated identity: Relevant vs. Irrelevant) × 3 (goal: Equality/Control/Loyalty) ANOVA. A significant interaction effect emerged, $F(2, 60) = 4.53$, $p = .015$, $\eta^2_p = .13$. No other effect approached significance. This interaction was investigated by analyzing the effect of the activated goal, separately in the relevant and irrelevant identity conditions.

The one-way ANOVA on *d'* indexes from the irrelevant identity conditions was not significant, $F(2, 30) < 1$, *ns*, indicating that the level of automatic favoritism was not influenced by the activated goal. On the contrary, results from the relevant identity conditions revealed a main effect of the activated goal, $F(2, 30) = 5.71$, $p = .008$, $\eta^2 = .18$. A linear trend emerged, $F(1, 30) = 10.28$, $p = .003$, with the lowest degree of intergroup bias in the equality condition, the highest degree of bias in the loyalty condition, and the control condition falling in between. Mean values of the GNAT index in both identity conditions are reported in Table 1.

In short, results from the relevant identity conditions replicate the influence of goals of equality and loyalty on implicit favoritism, initially demonstrated by Experiment 1. In these conditions, a significant linear trend confirmed that the lowest amount of implicit favoritism emerged after equality priming, and the highest amount of favoritism after loyalty priming,

whereas the control condition fell in between. Furthermore, results show a boundary condition for the effects of loyalty and equality: these goals exerted their influence on the implicit expression of evaluative bias only when no change in the intergroup context intervened between goal activation and attitude expression. In line with our prediction, the previous activation of an irrelevant identity acted as a boundary condition for the influence of equality and loyalty.

Experiment 3

Experiments 1 and 2 revealed how the activation of equality and loyalty goals influences implicit favoritism. An important question emerges from these findings and was addressed in the present experiment: would this goal activation exert a parallel influence on people's behavior? As discussed in the introductory section, we considered a spontaneous behavior characterized by ecological validity and by a low level of control, namely seating distance.

Additional aims of the present study were to further replicate the effects observed in Experiments 1 and 2 within a religious intergroup context, and to investigate whether goal activation influenced the level of implicit ingroup identification and explicit reports of attitudes toward target groups. In relation to implicit ingroup identification, it was conceivable that equality concerns decrease the importance of ingroup membership, because equality requires disregarding memberships in one's decision. Loyalty concerns, by contrast, might increase the importance of membership, because being loyal to one's own group requires a clear appreciation of one's membership. In line with this reasoning, Hertel and Kerr (2001) observed that priming of loyalty led to an increased expression of explicit ingroup identification, as compared with an equality condition, in a minimal group paradigm. By contrast, we did not expect any effect of equality or loyalty to emerge on participants' responses to direct questions, because of their highly controlled nature, combined with the strong presence of conscious normative pressures to repress any bias.

In Experiment 3 (as in Experiment 1) a goal was primed in the absence of any salient intergroup context. The experiment consisted of two phases that took place in separate laboratories. In the first phase, participants completed (in order) the goal activation task, a GNAT for the assessment of ingroup identification, an IAT for implicit favoritism, and two explicit questions on intergroup attitudes. In the second phase, participants were introduced to a get-acquainted task, in which the measure of seating distance was taken.

Method

Participants Sixty students (30 males, 30 females) took part to this experiment on a voluntary basis. They were approached by a female experimenter in the psychology building and asked to take part in an experiment on social cognition. All of them were Catholics.

Materials

Scrambled-Sentence Task The same scrambled-sentence task was used for goal activation as in Experiment 2.

Go/No-Go Association Task Six pictures related to Catholicism (e.g. Jesus, the Nativity), six pictures related to Islamism (e.g. the Hilal symbol, a mosque), six pronouns related to the self (e.g. I, me, myself), and six pronouns related to others (e.g. they, themselves, others) served as materials. The GNAT consisted of a series of five blocks of trials. The structure of each trial was the same as in Experiment 2. The first series of three blocks allowed participants to learn the response mode for the following categories: self, others, Catholics. The remaining two blocks were critical. In each critical block, participants were requested to recognize members of two target categories. In one critical block the target categories were Catholics and self. In the other, the categories were Catholics and others. Participants were presented with 32 trials in the learning blocks and with 90 trials in the critical blocks. The order of the blocks was randomized within the first series of three learning blocks and within the second series of two critical blocks.

Implicit Association Test Stimuli for the IAT were three pictures related to Catholics and three pictures related to Muslims which were different from those used in the GNAT, together with six positive and six negative adjectives (e.g. beautiful, nice, hostile, undesirable). The structure of the IAT was similar to Experiment 1, with the exceptions that: (i) images instead of words were used to represent the concept (i.e. Catholics, Muslims); (ii) stimuli referring to the concept and to the attribute were alternated; and (iii) only one order of blocks was presented, in which the congruent block associating the category of Catholics and positive words (and the category of Muslims and negative words) was presented before the other critical block. Participants were presented with 20 trials in the learning blocks, and 40 trials in the critical blocks.

Procedure Participants took part in the experiment in individual sessions and were randomly assigned to experimental conditions, with the limitation that approximately equal numbers of male and female respondents were assigned to each experimental condition. They were first administered the scrambled-sentence task, in one of three versions according to the condition (equality, loyalty, control). Subsequently, after a three-minute filler task aimed at clearing their working memory, they were seated in front of a computer and performed the GNAT and IAT, in this order. Then, two questions appeared on separate screens of the computer, one after the other: 'Please describe your attitude toward Catholics [Muslims].' Responses were provided on two 7-point Likert scales anchored at 1 (*extremely negative*) and 7 (*extremely positive*). Subsequently, participants were informed that they were going to become acquainted with a young Muslim man who was waiting in another room. The experimenter accompanied participants to the other room, but nobody was there. On a chair near the table there were a jacket and an Arabic-language newspaper, apparently belonging to the young Muslim man. Besides the chair occupied by these items, there was a pile of chairs in a corner. The experimenter commented that the other person was probably drinking a coffee in a room nearby, and suggested that the participant take a seat and fill in the last questionnaire while waiting for him. The participant took a chair from the pile, placed it near the table, sat down and filled in the questionnaire, which contained biographical questions (e.g. gender, religious affiliation). At this point, the experimenter informed them that the experiment was over, thanked them for participation and fully debriefed them. Participants were not explicitly probed for suspicion, but the experimenter was instructed to pay careful attention to any indication that participants may have doubted about the presence of the Muslim person. The experimenter reported that no participant expressed any suspicion, and all seemed very surprised when they were informed that the experiment was over.

Results and discussion

Preliminary analyses indicated that gender of participants had no significant effect. Hence, this factor was discarded from subsequent analyses.

Implicit ingroup identification (GNAT) For each of the two critical blocks, signal detection analysis was performed for each participant on the numbers of correct reactions and false alarms, separately for pronouns and for pictures. Beta indexes were submitted to a 3 (activated goal: Loyalty/Equality/Control) × 2 (type of stimuli: Picture/Word) × 2 (block: Self-Catholic vs. Others-Catholic) mixed model ANOVA, with the first factor manipulated between participants. A main effect of type of stimuli emerged, $F(1, 57) = 67.67$, $p < .001$, $\eta^2_p = .54$, indicating that participants had adopted a lower criterion for reacting to words ($M = .96$), as compared with pictures ($M = 2.07$); no other effect reached significance (all p values $> .10$). Hence, participants were more careful in reacting to pictures, as compared with pronouns, but the criterion did not change across experimental conditions. Therefore, we separately considered the indexes of discriminability for words and pictures. To calculate the index of identification with Catholics, we first computed: (i) the difference between the d' index of sensitivity for the detection of pictures referred to Catholics in

the task associating 'self' and Catholics, and in the task associating 'others' and Catholics; and (ii) an analogous difference for the detection of words. Our final index was the average of these two differences. The correlation of $r = .617$, $p < .001$ between the two d' scores for the detection of pictures and words indicated a satisfactory level of internal consistency for the GNAT.

Because 28% of our sample had not received the sacrament of Confirmation, which is an affirmation of the Christian faith, we could perform a known-group validation of the implicit identification index, by comparing its mean value for those who had and those who had not received Confirmation. As expected, those who had received Confirmation displayed significantly higher levels of identification, as compared with those who had not, $t(58) = 2.10$, $p < .05$, $d = .74$; mean values and standard deviations were $M = .03$ and $SD = .91$, and $M = -.53$ and $SD = .58$, for confirmed and non-confirmed participants respectively. No effect of the activated goal on implicit identification of participants emerged, $F(2, 57) < 1$, *ns*, $\eta^2_p = .03$. Hence, the GNAT index proved to be a valid and sensitive measure of implicit ingroup identification, and was not influenced by the primed goal. Mean levels of implicit identification with the ingroup, depending on the experimental condition, are reported in Table 1.

Implicit ingroup favoritism (IAT)

An index of implicit ingroup favoritism was computed from the latencies of responses to the critical blocks of the IAT, following Greenwald et al. (1998). After discarding extremely fast (< 150 ms) and extremely slow responses (> 10,000 ms), latencies were trimmed to fall in the 300–3000 ms interval. We subsequently computed for each participant the mean latency for the Catholic-positive and for the Catholic-negative block. The index of implicit ingroup favoritism was computed subtracting the first from the second mean latency value.[5] The IAT had an acceptable internal consistency, $\alpha = .74$.

A one-way ANOVA compared the mean level of implicit favoritism across experimental conditions. A significant effect emerged, $F(2, 57) = 3.33$, $p < .05$, $\eta^2 = .10$. A significant linear trend indicated that the display of intergroup bias was highest in the loyalty condition and lowest in the equality condition, with the control condition falling in between, $F(1, 57) = 6.14$, $p < .02$. Average levels of intergroup bias, according to the experimental condition, are reported in Table 1.

Explicit intergroup bias A two-way ANOVA on the explicit attitude toward Catholics and Muslims, with target group as a within participant variable, and activated goal as a between participants variable, revealed no significant main or interaction effect, all F values < 1.60, all p values > .20. In other words, participants did not display explicit favoritism, and they were not influenced by the experimental condition in their explicitly expressed evaluations.

Seating distance A one-way ANOVA on seating distance revealed that the activated goal had a significant effect, $F(2, 57) = 4.57$, $p = .014$, $\eta^2 = .14$. A clear linear trend indicated that participants sat nearer to the chair of the Muslim when the equality goal was activated, and farther away when the loyalty goal was activated, as compared with the control condition, in which no goal was activated, $F(1, 57) = 9.05$, $p < .005$. Mean levels of seating distance, depending on the experimental condition, are reported in Table 1. In this study, the IAT and seating distance measures did not correlate with each other, $r(58) = .09$, *ns*. In this respect, we failed to replicate some past research that had found them to be related (Amodio & Devine, 2006). However, our priming manipulation influenced both of these variables even though they did not share significant variance with each other, indicating its potency in affecting both implicit attitudes and behavior.

The present results are congruent with Hertel and Kerr's (2001) study, which showed that equality and loyalty priming influenced resource allocation behavior, but an interesting dissimilarity emerged in ingroup identification. Hertel and Kerr's results showed a significant enhancement in the expression of ingroup identification of participants after loyalty priming, as compared with the equality condition. No influence of priming on identification emerged

in the present study. A series of important differences between the present study and Hertel and Kerr's experiment may help explain this apparent incongruence. First, we used a measure of implicit identification, whereas Hertel and Kerr used a questionnaire measuring explicit identification. It may be that priming loyalty did not influence true feelings of ingroup identification, but instead increased the explicit statement of ingroup identification, which may be considered as an ingroup supporting activity. Second, it is possible that the bipolar priming technique (Hertel & Fiedler, 1994) utilized by Hertel and Kerr (2001), by stressing the negative evaluation of norm defiance, caused stronger or slightly different effects, as compared to our priming manipulation. Most importantly, Hertel and Kerr's study was conducted within a minimal group paradigm. The meaning of identification may be quite different in real and in minimal groups. Given the number of differences between our studies and Hertel and Kerr's (2001) study, it is difficult to point to the aspect of the procedures that caused the observed differences in results.

To summarize the main results from the Experiment 3, the activated goals of equality and loyalty had no influence on levels of implicit identification or on explicit attitudes toward ingroup and outgroup, but, congruent with our hypotheses, they influenced both implicit favoritism and spontaneous behavior. Indeed, replicating results from previous studies, participants primed with loyalty showed higher levels, and participants primed with equality showed lower levels of bias, as compared with participants in the control condition. Most importantly, activated goals influenced seating distance: participants primed with loyalty sat farther away, and participants primed with equality sat nearer to the chair occupied by the member of a stigmatized group, as compared with participants in the control condition.

Meta-analytic summary of the studies

In the three studies presented in this article, we used different measures of intergroup bias (IAT, GNAT, and seating distance) to investigate the impact of temporary goals of equality and loyalty. As expected, in all studies and in all measures a significant linear trend consistently emerged, indicating that the level of intergroup bias expressed after equality activation was minimal, compared with an intermediate level in the control condition and a high level after activation of loyalty. Because of the low sample sizes and the consequent limited power of pairwise comparisons in each study, to summarize the effects of loyalty and equality on the expression of intergroup bias, we statistically combined the results of the three studies.

As a first step, two separate effect sizes were calculated for each measure: the impact of equality activation compared with the control condition, and to the impact of loyalty activation, also compared with the control. Each effect size was the standardized difference between mean responses in the experimental and in the control condition. Effect sizes were computed as d (Cohen, 1988); a positive effect size indicates a hypothesis-consistent result, and a negative effect size indicates a hypothesis-inconsistent result. We computed eight different effect sizes, assessing the effects of equality and loyalty for the IAT index of Experiment 1, the GNAT index of Experiment 2, the IAT index of Experiment 3, and the seating distance measure of Experiment 3.

Because in Experiment 3 two different measures of intergroup bias were collected, we summarized their respective effect sizes separately for the two experimental conditions, and weighted them according to the inverse to their variance (see Cooper & Hedges, 1994). Finally, the overall effect size of the meta-analysis of the three presented studies was calculated using the weighted integration method based on the inverse of variance. In Table 2, effect sizes are reported for each study together with summary indexes. As can be seen in Table 2, an effect size of $d = .50$ emerged for the comparison between equality and control conditions, with a 95% confidence interval ranging from .19 to .81. For the comparison between loyalty and control conditions, a mean effect size of $d = .38$ emerged, with a 95% confidence interval ranging from .06 to .67. Both these overall

Table 2. Summary of results from the three experiments

		Statistics for each study		95% confidence interval			
Experiment	Dependent measure	*d*	SE	Lower limit	Upper limit	*Z*	*p*
Effect of equality priming							
1	IAT	0.26	0.25	–0.24	0.75	1.02	.156
2	GNAT	1.22	0.46	0.31	2.14	2.64	.004
3	IAT	0.65	0.32	0.02	1.29	2.01	.023
	Seating distance	0.41	0.32	–0.22	1.04	1.29	.099
	Weighted effect size	0.53	0.23	0.08	0.98	2.32	.010
Combined studies		0.5007	0.16	0.19	0.81	3.16	.001
Effect of loyalty priming							
1	IAT	0.37	0.25	–0.12	0.87	1.48	.069
2	GNAT	0.29	0.43	–0.55	1.13	0.69	.248
3	IAT	0.18	0.32	–0.44	0.80	0.58	.281
	Seating distance	0.59	0.32	–0.04	1.23	1.83	.034
	Weighted effect size	0.38	0.23	–0.06	0.83	1.70	.045
Combined studies		0.37	0.16	0.06	0.67	2.34	.010

effect sizes are statistically significant, $p < .001$ and Rosenthal's (1979) fail-safe $N_R = 10.05$ for the effect of equality; $p < .01$, and Rosenthal's (1979) fail-safe $N_R = 2.44$ for the effect of loyalty (desired $\alpha = .05$).

It is important to note that both equality and loyalty priming exerted a significant influence on intergroup bias, which can be classified as small-to-medium according to Cohen (1988). The significant impact of loyalty priming on intergroup bias demonstrates that, although considerable implicit favoritism is found in the control condition, this may be further enhanced by the context. Because of the small set of studies on which they are based, we believe that no clear conclusion can be drawn from the differences in the overall effect sizes for equality and loyalty priming. The safest conclusion seems to be that using different methodologies and target groups, we found small to medium but reliable effects both of equality and of loyalty activation on automatic evaluative bias.

General discussion

The present research investigated how contextually primed intergroup goals influence implicit favoritism and spontaneous intergroup behavior. It was posited that the indirect activation of equality would attenuate the expression of intergroup bias, and the indirect activation of loyalty would enhance its expression. Experiment 1 supported these hypotheses.

The results underline the importance of the contextual modulation of implicit favoritism. Previous studies provided important clues on this issue (Barden et al., 2004; Dasgupta & Greenwald, 2001; Sinclair et al., 2005). Here, we show that the activation of equality and loyalty may also shape spontaneous responses. Indeed, the mere activation of these goals through an alleged language test influences both implicit evaluative bias and corresponding behavior.

A second important result from Experiment 1, confirmed by the subsequent experiments, is the finding that implicit favoritism, although it is easily detected and reaches substantial levels in ordinary conditions, is not generally at its maximum level. As a result, there is room for exacerbation of its virulence, given the appropriate conditions. As indicated in the introduction, no study provided this result before.

In Experiment 2, it was further shown that boundary conditions are present for the observed results. In particular, no influence of the activated equality and loyalty goals emerged when an irrelevant intergroup context was made salient before goal activation. According to an interpretation of representations as context-based states (Smith & Conrey, 2007), this

result suggests that equality and loyalty, despite being very general and broad values, can take a specific context-dependent meaning. Thereafter, they exert their influence only when they are applicable, in their specific connotation, to the task at hand.

In Experiment 3, another important step was accomplished, by showing that intergroup behavior can be influenced by this contextual manipulation. In particular, participants sat farther away from an outgroup member after the activation of loyalty goals, and closer to this person after the activation of equality goals, with the control condition falling in between.

Temporarily activated goals and implicit ingroup favoritism

Recent research clearly indicates that goal pursuit may be triggered by the environment and guided by unconscious processes. Various social goals, like cooperation or socialization, may be activated by environmental stimuli and subsequently guide thought and behavior (Aarts et al., 2005; Aarts, Gollwitzer, & Hassin, 2004; Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trötschel, 2001; Chartrand & Bargh, 1996; Custers & Aarts, 2005; Fishbach, Friedman, & Kruglanski, 2003; Shah, 2003), and the present study extends this literature by showing the importance of equality and loyalty goals for social perception. Previous research showed that the processing of information related to a certain goal produces motivational effects, like inhibition of competing goals and enhancement of effort toward goal pursuit (see Moskowitz et al., 2004). The psychological model underlying this line of research considers goal representations as similar to other concepts, which may be triggered directly by the environment, or indirectly through spreading activation from associated nodes.

In the present set of studies, there are several indications that participants were unaware of the influence of priming on their expressions of intergroup bias, and therefore that implicit priming effects were observed (Bargh, 1992). The indirect activation of goals can produce results that are analogous to those caused by explicit directions. For example, subliminal activation of cooperative goals produced an enhancement in cooperative behavior that was similar to that produced by overt instructions to cooperate, despite participants' lack of consciousness of goal activation (Bargh et al., 2001; see also Chartrand & Bargh, 1996). However, there are differences between implicit and explicit goal activation, which are potentially important in the present case. In the first place, when people are conscious of environmental requests, they may decide to comply and behave accordingly, but reactance effects related to the need to preserve one's psychological autonomy could take place (Brehm, 1966). Aside from reactance, effects of explicit instructions may be minimal when people disagree with the request to adopt a certain goal (see Plant & Devine, 2001). In both cases, either because of reactance for the protection of one's autonomy, or because of disagreement with the social goal, explicit instructions might have no effect or even produce a backlash. Even though this hypothesis was not tested in the present set of studies, it is possible that the indirect mode of goal activation was an important condition for the emergence of moderating effects of the social goals on intergroup bias.

Interpretation of the observed changes in implicit attitude expression

The present results are highly congruent with the justification suppression model (JSM). The JSM posits that genuine prejudice consists in 'pure, unadulterated [. . .] negative feelings toward members of devalued groups' (Crandall & Eshleman, 2003, p. 418). JSM distinguishes two structural elements that alter the expression of prejudice: suppression and justification. Suppression is defined as any motivated attempt to reduce the expression or awareness of prejudice. These attempts are thought to lead to a more favorable outward attitude, without any change in genuine prejudice. The value of equality is listed by Crandall and Eshleman (2003) among the sources of suppression. Justification, by contrast, is secondary to suppression because it acts as a releaser of otherwise suppressed genuine prejudices. It is defined as any social or psychological process that gives the opportunity

to express one's genuine prejudice without undergoing external or internal sanction. Albeit not specifically listed by Crandall and Eshleman (2003), loyalty may be conceived as a factor justifying favoritism toward one's ingroup. Importantly, according to JSM, the observed influences of equality and loyalty should be not interpreted as changes in genuine intergroup attitudes, but only as changes in attitude expression. If we accept this interpretation of our results, one novel point is that these shifts in expression occur (i) with participants who are unaware of the connection between the priming task and the dependent measures, and (ii) on implicit dependent measures. In other words, shifts in attitude expression in these studies cannot be regarded as conscious and strategic in nature.

Two different accounts, however, would be compatible with the notion that activated goals influenced participants' genuine attitudes in the present studies rather than simply tuning their expression. One explanation involves changes in the level of categorization, another is an interpretation in terms of context-dependent representations.

According to an explanation in terms of changes in the level of categorization and attention devoted to intergroup differentiation, it is plausible that equality priming could distract attention from the relevant intergroup differentiation, either by enhancing individuation (Brewer, Weber, & Carini, 1995), or by causing categorization at a more comprehensive level of common ingroup identity (Gaertner & Dovidio, 2005). Loyalty, by contrast, could enhance attention to the salient intergroup differentiation. Differences in the degree of spontaneous categorization at the relevant intergroup level should, in turn, influence intergroup bias (see Brown & Hewstone, 2005; Gaertner & Dovidio, 2005; Hornsey & Hogg, 2000). Support for this explanation was provided in a study on spontaneous categorization (Zogmaister, 2004; Zogmaister et al., 2005), which revealed that the level of categorization of ingroup members in a 'Who said what' paradigm was minimum when equality was primed, and maximum when loyalty was primed, with the control condition falling in between.

A second theoretical possibility is in line with the notion of mental representations as context-sensitive and reconstructed states (Smith & Conrey, 2007). According to this view, attitude malleability could indicate that the same input (e.g. an ingroup) elicits different reconstructed representations, due to the effects of other aspects of the overall situation. The present results can be interpreted as evidence that motivational states, and in particular the goal to be equal or loyal, influence the way in which the intergroup context is represented by the cognitive system. This in turn influences the way in which ingroup and outgroup are evaluated. According to this perspective, implicit attitudes expressed by participants in all experimental conditions, inasmuch as they are not intentionally and consciously tuned to convey a specific image to the experimenter, may be considered genuine reactions to the contextual representations of the groups.

Conclusions

Our main interest in the present set of studies was to understand how to make use of social goals in order to 'put the brakes on prejudice'. We believe it is important both to investigate approaches that decrease prejudice expression (in order to support their use), and to understand factors that enhance prejudice (in order to avoid their presence, especially in potentially conflict-loaded intergroup contexts). In the present paper it was shown that social goals are important both for the prevention of prejudice, and for the promotion of intergroup justice.

Notes

1. Parallel analyses were conducted on the less sensitive *D* index, proposed by Greenwald, Nosek and Banaji (2003). The pattern of results emerging from these analysis was the same as described here, but the linear trend was only marginally significant, $F(1, 93) = 3.576$, $p = .060$, equality–control contrast $t(62) = 1.47$, $p = .14$ (two-tailed), control–loyalty contrast $t(62) < 1$, *ns*.

2. Previous studies showed that ingroup identities and intergroup attitudes of Southern and Northern Italians do not simply mirror each other (Sanchez et al., 2007; Zogmaister, Arcuri, & Modena, 2006). Because we could not collect data from a sufficiently numerous sample of Southern respondents, our analysis must be restricted to Northern Italians.
3. In two cases participants failed to turn off their cell phone and it rang during the experimental session, in one case the participant failed to complete the identity activation task, and in five cases participants failed to comply with GNAT instructions.
4. A further reason for preferring, in our main analyses, the indexes of discriminability based on ingroup/outgroup discrimination is theoretically based. Because we are interested in the valence that is automatically associated with a specific target, depending on the experimental condition, differential reactions to the target, depending on the pole of the evaluative dimension it is associated with, are more informative for our ends, as compared with reactions to the evaluative dimensions themselves.
5. As in Experiment 1, a second index of implicit bias in the IAT was computed following Greenwald et al. (2003). The two indexes were highly correlated, $r_{60} = .931$, $p < .001$. The same pattern of results emerged with this second index, as that presented in the paper. These analyses are available from the first author on request.

References

Aarts, H., Chartrand, T. L., Custers, R., Danner, U., Dik, G., Jefferis, V. E. et al. (2005). Social stereotypes and automatic goal pursuit. *Social Cognition, 23*, 465–490.

Aarts, H., Gollwitzer, P., & Hassin, R. (2004). Goal contagion: Perceiving is for pursuing. *Journal of Personality and Social Psychology, 87*, 23–37.

Abrams, D., Marques, J., Bown, N., & Dougill, M. (2002). Anti-norm and pro-norm deviance in the bank and on the campus: Two experiments on subjective group dynamics. *Group Processes & Intergroup Relations, 5*, 163–182.

Abrams, D., Marques, J., Bown, N. J. & Henson, M. (2000). Pro-norm and anti-norm deviance within and between groups. *Journal of Personality and Social Psychology, 78*, 906–912.

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology, 91*, 652–661.

Argyle, M., & Dean, J. (1965). Eye contact, distance, and affiliation. *Sociometry, 28*, 289–304.

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology, 87*, 5–22.

Bargh, J. A. (1992). Does subliminality matter to social psychology? Awareness of the stimulus versus awareness of its influence. In R. F. Bornstein & T. S. Pittman (Eds.), *Perception without awareness* (pp. 236–255). New York: Guilford Press.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automatized will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology, 81*, 1014–1027.

Barrios, B. A., Corbitt, L. C., Estes, J. P., & Topping, J. S. (1976). Effect of a *social* stigma on interpersonal *distance. Psychological Record, 26*, 343–348.

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

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*, 828–841.

Blank, T., & Schmidt, P. (2003). National identity in a united Germany: Nationalism or patriotism? An empirical test with representative data. *Political Psychology, 24*, 289–311.

Braithwaite, V. A., & Law, H. G. (1985). Structure of human values: Testing the adequacy of the Rokeach Value Survey. *Journal of Personality and Social Psychology, 49*, 250–263.

Brehm, J. W. (1966). *A theory of psychological reactance.* New York: Academic Press.

Brewer, M. B., Weber, J. G., & Carini, B. (1995). Person memory in intergroup contexts: Categorization versus individuation. *Journal of Personality and Social Psychology, 69*, 29–40.

Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 37, pp. 255–343). San Diego, CA: Academic Press.

Byrne, D. (1961). Interpersonal attraction and attitude similarity. *Journal of Abnormal Social Psychology, 62*, 713–715.

Castelli, L., Macrae, N. C., Zogmaister, C., & Arcuri, L. (2004). A tale of two primes: Contextual limits on stereotype activation. *Social Cognition, 22*, 233–247.

Castelli, L., Tomelleri, S., & Zogmaister, C. (2008). Implicit ingroup metafavoritism: Subtle preference for ingroup members displaying ingroup bias. *Personality and Social Psychology Bulletin, 34*, 807–818.

Chartrand, T. L., & Bargh, J. A. (1996). Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effects of explicit task instructions. *Journal of Personality and Social Psychology, 71*, 464–478.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conrey, F. R., & Smith, E. R. (2007). Attitude representation: Attitudes as patterns in a distributed, connectionist representational system. *Social Cognition, 25*, 718–736.

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Crace, R. K., & Brown, D. (1996). *Life values inventory*. Chapel Hill, NC: Life Values Resources.

Crandall, C. S., & Eshleman, A. (2003). A justification–suppression model of the expression and experience of prejudice. *Psychological Bulletin, 129*, 414–446.

Custers, R., & Aarts, H. (2005). Positive affect as implicit motivator: On the nonconscious operation of behavioral goals. *Journal of Personality and Social Psychology, 89*, 129–142.

Dambrun, M. L., & Guimond, S. (2004). Implicit and explicit measures of prejudice and stereotyping: Do they assess the same underlying knowledge structure? *European Journal of Social Psychology, 34*, 663–676.

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*, 800–814.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62–68.

Dovidio, J., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510–540.

Fishbach, A., Friedman, R. S., & Kruglanski, A. W. (2003). Leading us not unto temptation: Momentary allurements elicit overriding goal activation. *Journal of Personality and Social Psychology, 84*, 296–309.

Gaertner, S. L., & Dovidio, J. F. (2005). Understanding and addressing contemporary racism: From aversive racism to the common ingroup identity model. *Journal of Social Issues, 61*, 615–639.

Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation and retrieval experiences. *Journal of Personality and Social Psychology, 89*, 672–685.

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When 'Just Say No' is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology, 44*, 370–377.

Gollwitzer, P. M., & Schaal, B. (1998). Metacognition in action: The importance of implementation intentions. *Personality and Social Psychology Review, 2*, 124–136.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: 1. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197–216.

Hendricks, M., & Bootzin, R. (1976). Race and sex as stimuli for negative affect and physical avoidance. *Journal of Social Psychology, 98*, 111–121.

Hertel, G., & Fiedler, K. (1994). Affective and cognitive influences in a social dilemma game. *European Journal of Social Psychology, 24*, 131–145.

Hertel, G., & Kerr, N. L. (2001). Priming in-group favoritism: The impact of normative scripts in the minimal group paradigm. *Journal of Experimental Social Psychology, 37*, 316–324.

Hornsey, M. J., & Hogg, M. A. (2000). Subgroup relations: A comparison of mutual intergroup differentiation and common ingroup identity models of prejudice reduction. *Personality and Social Psychology Bulletin, 26*, 242–256.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology, 78*, 871–888.

Kühnen, U., Schiessl, M., Bauer, N., Paulig, N., Poehlmann, C., & Schmidthals, K. (2001). How robust is the IAT? Measuring and manipulating implicit attitudes of East- and West-Germans. *Experimental Psychology, 48*, 135–144.

Levine, J. M., & Moreland, R. L. (2002). Group reactions to loyalty and disloyalty. In S. R. Thye & E. J. Lawler (Eds.), *Group cohesion, trust and solidarity: Advances in group processes* (Vol. 19, pp. 203–228). Oxford, UK: Elsevier Science.

Li, Q., & Brewer, M. B. (2004). What does it mean to be an American? Patriotism, nationalism, and American identity after 9/11. *Political Psychology, 25*, 727–739.

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology, 69*, 397–407.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology, 67*, 808–817.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., Thorn, T. M. J., & Castelli, L. (1997). On the activation of social stereotypes: The moderating role of processing objectives. *Journal of Experimental Social Psychology, 33*, 471–489.

Macrae, C. N., Mitchell, J. P., & Pendry, L. F. (2002). What's in a forename? Cue familiarity and stereotypical thinking. *Journal of Experimental Social Psychology, 38*, 186–193.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77*, 167–184.

Moskowitz, G. B., Li, P., & Kirk, E. (2004). The implicit volition model: On the preconscious regulation of temporarily adopted goals. In M. P. Zanna (Ed.), *Advances in experimental social psychology, Vol. 36* (pp. 317–413). San Diego, CA: Academic Press.

Moskowitz, G. B., Salomon, A. R., & Taylor, C. M. (2000). Preconsciously controlling stereotyping: Implicitly activated egalitarian goals prevent the activation of stereotypes. *Social Cognition, 18*, 151–177.

Nosek, B., & Banaji, M. (2001). The Go/No-Go Association Task. *Social Cognition, 19*, 625–666.

Park, B., & Judd, C. M. (2005). Rethinking the link between categorization and prejudice within the social cognition perspective. *Personality and Social Psychology Review, 9*, 108–130.

Plant, E. A., & Devine, P. G. (2001). Responses to other-imposed pro-Black pressure: Acceptance or backlash? *Journal of Experimental Social Psychology, 37*, 486–501.

Richeson, J. A., & Ambady, N. (2003). Effects of situational power on automatic racial prejudice. *Journal of Experimental Social Psychology, 39*, 177–183.

Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology, 40*, 417–423.

Rokeach, M. (1967). *Value survey*. Sunnyvale, CA: Halgren Tests.

Rokeach, M. (1973). *The nature of human values*. New York: Free Press.

Rosenthal, R. (1979). The 'file drawer problem' and tolerance for null results. *Psychological Bulletin, 86*, 638–641.

Sanchez, A. K., Zogmaister, C., & Arcuri, L. (2007). When 'They' becomes 'We': Multiple contrasting identities in mixed status groups. *Self and Identity, 6*, 154–173.

Sassenberg, K., & Moskowitz, G. B. (2005). Don't stereotype, think different! Overcoming automatic stereotype activation by mindset priming. *Journal of Experimental Social Psychology, 41*, 506–514.

Scott, W. A. (1965). *Values and organizations: A study on fraternities and sororities*. Chicago: Rand McNally.

Sechrist, G. B., & Stangor, C. (2001). Perceived consensus influences intergroup behavior and stereotype accessibility. *Journal of Personality and Social Psychology, 80*, 645–654.

Shah, J. (2003). Automatic for the people: How representations of significant others implicitly affect goal pursuit. *Journal of Personality and Social Psychology, 84*, 661–681.

Sinclair, L., & Kunda, Z. (1999). Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77*, 885–904.

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*, 583–592.

Smith, E. R., & Conrey, F. R. (2007). Mental representations are states, not things: Implications for implicit and explicit measurement. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 247–264). New York: Guilford Press.

Spencer, S. J., Fein, S., Wolfe, C., Fong, C., & Dunn, M. (1998). Stereotype activation under cognitive load: The moderating role of self-image threat. *Personality and Social Psychology Bulletin*, *24*, 1139–1152.

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, *37*, 1660–1672.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247.

Swets, J. A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin*, *99*, 181–198.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. Cambridge, UK: Cambridge University Press.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago: Nelson-Hall.

Van Vugt, M., & Chang, K. (2006). Group reactions to loyal and disloyal high status members. Unpublished manuscript.

Van Vugt, M., & Hart, C. M. (2004). Social identity as social glue: The origins of group loyalty (2004). *Journal of Personality and Social Psychology*, *86*, 585–598.

Zdaniuk, B., & Levine, J. M. (2001). Group loyalty: Impact of members' identification and contributions. *Journal of Experimental Social Psychology*, *37*, 502–509.

Zogmaister, C. (2004). *Social norms and intergroup bias*. Doctoral dissertation, University of Padova, Italy.

Zogmaister, C., Arcuri, L., Castelli, L., Smith, E. R., Paladino, M. P., & Seger, C. (2005, July). *Social norms and intergroup bias*. Communication presented at the 14th General Meeting of the European Association of Experimental Social Psychology, Würzburg, Germany.

Zogmaister, C., Arcuri, L., & Modena, S. (2006). La percezione di gruppi caratterizzati da status sociale asimmetrico. Rapporto tra processi impliciti ed espliciti [The perception of groups characterized by asymmetrical status. The relationship between implicit and explicit processes]. *Rassegna di Psicologia*, *13*, 93–122.

Paper received 6 June 2007; revised version accepted 1 April 2008.

Biographical notes

CRISTINA ZOGMAISTER is assistant professor at the University of Milan-Bicocca (Italy). She received her PhD in cognitive sciences in 2005. Her main research interests are in implicit social cognition, and more specifically on the influences of social context on spontaneous judgments and behaviors, with a particular attention on implicit measures of social-cognitive constructs.

LUCIANO ARCURI is professor of social psychology at the University of Padova (Italy). He works in the area of social cognition. Over recent decades he has conducted experimental research on stereotyping and intergroup relations, in particular using techniques involving implicit measures. Recently he has investigated automatic processes involved in the activation and suppression of stereotypes.

LUIGI CASTELLI is associate professor of social psychology at the University of Padova (Italy). His research interests are mainly focused on implicit intergroup perception and the development of prejudice among children.

ELIOT R. SMITH is professor of psychological and brain sciences at Indiana University, Bloomington. His research interests include stereotyping and prejudice, with a special focus on the role of emotions, and interfaces between social psychology and situated/embodied cognition.