British Journal of Social Psychology (2010), 49, 43–67

The British Psychological Society

www.bpsjournals.co.uk

To be or not to be: The impact of implicit versus explicit inappropriate social categorizations on the self

Manuela Barreto[1]*, Naomi Ellemers[2], Wieke Scholten[2] and Heather Smith[3]

[1]Centre for Social Research and Intervention, Lisbon, Portugal
[2]Leiden University, Institute for Psychological Research (LU-IPR), Leiden, The Netherlands
[3]Sonoma State University, Rohnert Park, California, USA

This paper investigates how targets respond to treatment that is explicitly or implicitly based on a *contextually inappropriate* social categorization. In three different experimental studies, a team member appeared to use participants' gender and not participants' personal preference to assign a proofreading task. Targets reported more negative self-evaluations in response to implicit categorical treatment in comparison to explicit categorical treatment. In contrast, explicit categorical treatment increased target's resistance to the treatment received. The pattern of results across the three studies shows that treatment based on a contextually inappropriate category is problematic even when the categorization is ambiguous or associated with attractive and positive outcomes.

Imagine that it is Sunday morning and you are attending your child's football match. Now imagine that you also happen to be a medical doctor and the parent standing by your side, knowing this, takes the opportunity to quiz you about the best way to manage diabetes. Most of us can recount episodes in which we were treated on the basis of an identity that we would rather not see as self-defining in that particular context. The purpose of this paper is to explore the psychological consequences of such inappropriate categorizations when these categorizations are explicit and when they are less obvious.

Social psychologists long have demonstrated that people try to convince others to share their own image of themselves (e.g. Swann & Read, 1981). This process of self-verification appears to be so important that people try to convince others of their preferred self-image even when that self-image is negative (e.g. Swann, 1990). However,

*Correspondence should be addressed to Dr Manuela Barreto, Centre for Social Research and Intervention (CIS), Av das Forças Armadas–Ed ISCTE, 1649-026 Lisbon, Portugal (e-mail: manuela.barreto@iscte.pt).

DOI:10.1348/014466608X400830

despite our best efforts, others do not always share our own self-views, and we often are confronted with treatment that reveals this discrepancy. These discrepancies can be chronic – when people systematically base their views about us on visible features (e.g. gender) that are inconsistent with our preferred self-definition – or they may be more contextual – when people base their views about us on a category (e.g. occupation) that is inconsistent with our preferred self-view for a particular context (see also Barreto & Ellemers, 2003). The purpose of this paper is to examine how targets respond to treatment that is explicitly or implicitly based on a *contextually inappropriate* social categorization. In three different studies, a team member used participants' gender and not participants' personal preference to assign a proofreading task. Given the demonstrated importance and prevalence of self-verification processes, the use of contextually inappropriate categorizations is likely to have severe consequences for the target. We hypothesize that explicit inappropriate categorical treatment will lead to anger and protest whereas implicit categorical treatment will lead to negative self-views.

Multiple identities and categorization threat

When asked to indicate who they are, people describe themselves in many different ways. These self-descriptions can vary depending on the level of abstraction at which they are made – sometimes people see themselves as unique individuals whereas in other contexts they see themselves as group members (e.g. a faculty member; Deaux, 1996; Tajfel, 1981; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). People also may describe themselves as a member of a particular group in one context (e.g. a faculty member at work) and as a member of a different group in other contexts (e.g. a parent on the football field; Oakes, Haslam, & Turner, 1994).

Categorization threat occurs when other people use a social category to define a person that contradicts that person's contextually preferred self-definition. For example, gender is a highly visible cue and people often are categorized according to their gender. However, targets do not always define or categorize themselves according to the most visible categorization cues and this can lead to identity discrepancies (see also Barreto & Ellemers, 2003). These identity discrepancies can take different forms. For example, others most often categorize members of ethnic minorities based on their ethnic group membership (Horenczyk, 1996; Van Oudenhoven, Prins, & Buunk, 1998), but in many contexts ethnic minority employees may prefer to be treated as individuals, or even as members of the host society. In another instance, a female lawyer may prefer to be viewed on the basis of her professional identity at work, but at a party she may prefer to avoid being approached for legal advice and instead embrace being viewed as a woman.

Categorization threats are inherently contextual because the possible identities that can be imposed by others or preferred by the self depend upon the particular context. In one situation, a target may feel threatened when she is categorized as Moroccan instead of as a Dutch citizen. However, in a different situation, the same target may consider her Moroccan identity to be more relevant and feel threatened when she is categorized as Dutch.

Categorization threat, as we define it, is different from other types of identity threat (see also Barreto & Ellemers, 2003; Branscombe, Ellemers, Spears, & Doosje, 1999; Ellemers, Spears, & Doosje, 2002). First, categorization threat is not dependent on the existence of negative group stereotypes. Instead, the threat is caused by a mismatch between the category that others use and the category that we prefer, irrespective of the valence of the stereotypes associated with these categories. Because categorization

threat does not require an association with negative stereotypes, it is different from stereotype threat and from group-based discrimination.[1] For example, stereotype threat occurs when stigmatized group members fear that they might confirm negative stereotypes associated with their group (Steele & Aronson, 1995). Any environmental cue that makes a category membership salient in the context of a task for which the category member is expected to be inferior can reduce performance (see Schmader, Johns, & Forbes, 2008 and Steele, Spencer & Aronson, 2002 for recent reviews). In contrast, categorization threat captures situations in which people resent other people's imposition of quite positive group memberships (e.g. doctor) when they prefer a different self-definition (e.g. parent).

Group-based discrimination also requires some basis in negative stereotypes, even when only positive stereotypes are stressed. For example, benevolent sexism stresses positive stereotypes of women, but is rooted in both positive and negative stereotypes of women – as caring but also as incapable of taking care of themselves (Glick & Fiske, 1996). Also, discrimination is associated with negative outcomes. Even when discrimination appears to be benevolent (e.g. the positive affect or flattery associated with benevolent sexism), the ultimate outcomes are negative (lack of support for rape victims or victims of domestic violence, not being able to have a job, poor task performance, e.g. Abrams, Viki, Masser, & Bohner, 2003; Dardenne, Dumont, & Bollier, 2007; Glick, Sakalli-Ugurlu, Ferreira, & Souza, 2002). In contrast, it is not negative stereotypes or negative outcomes that create categorization threat – it is ignoring the person's preferred self-definition.

Second, categorization threats are different from group distinctiveness threats (see also Brewer, 1991; Hewstone & Brown, 1986). Distinctiveness and categorization threats can have similar effects: both can lead to explicit rejection of the imposed identity, to stressing the difference between the self and the imposed group (Spears, Doosje, & Ellemers, 1997), to lack of loyalty to the imposed group (Barreto & Ellemers, 2000; Jetten, Spears, & Manstead, 1997; Ouwerkerk, de Gilder, & de Vries, 2000), and to affirmation of alternative identities that are seen as more contextually appropriate (Barreto & Ellemers, 2002). However, these two types of threat are theoretically distinguishable. Indeed, although categorization threats might include threats that undermine a group's distinctiveness (i.e. when people wish to be treated on the basis of a distinctive identity but are not), this is not always the case. Categorization threats can arise when people are treated on the basis of a *more* distinctive identity but they prefer to be treated in terms of a less distinctive identity (e.g. an ethnic minority faculty member who prefers to be recognized for her occupational expertise).

Research shows that discrepancies between self-perceptions and externally imposed categorizations have negative consequences for intergroup relations. For example, immigrants who perceive a discrepancy between the way they see themselves and the way (they think that) members of the host society see them view the host society more negatively (Bourhis, Moise, Perreault, & Senecal, 1997). Similarly, if authorities treat subgroup members on the basis of a superordinate identity (and ignore their subgroup

[1] *Categorization threat is also different from positive discrimination. Positive discrimination is based on the existence of negative stereotypes about a particular group, even though it may be set in place to discourage those stereotypes. Moreover, categorization threat necessarily implies unwanted categorization, whereas the concept of positive discrimination does not involve any knowledge of whether or not the targets would like to be categorized as such. In fact, whether or not targets of positive discrimination want to be seen as members of their group is likely to be a matter of context and a matter of individual differences. Thus, targets of positive discrimination may or not experience categorization threat.*

membership), subgroup members express stronger subgroup identification and increased bias against other subgroups in comparison to subgroup members who feel that authorities acknowledge superordinate and subgroup identities (Gaertner, Rust, Dovidio, Bachman, & Anastasio, 1994; Haunschild, Moreland, & Murrell, 1994; Hornsey & Hogg, 2000; Huo & Molina, 2006; Terry, Carey, & Callan, 2001).

However, little is known about the impact of categorization threat *on the self*. Research on the affective consequences of discriminatory treatment (i.e. unwanted group-based treatment that involves in-group disadvantage) shows that exposure to prejudice and discrimination is often associated with negative affect and low personal self-esteem (e.g. Branscombe, Schmitt, & Harvey, 1999; Clark, Anderson, Clark, & Williams, 1999; Klonoff, Landrine, & Campbell, 2000). Also, research on procedural justice shows that whether or not people feel that their preferences are heard and taken into account within groups and organizations affects their personal self-esteem (Koper, Van Knippenberg, Bouhuijs, Vermunt, & Wilke, 1993; McAllister & Bigley, 2002; Pierce, Gardner, Cummings, & Dunham, 1989; Tyler, Degoey, & Smith, 1996). But, as far as we know, there is no evidence to show that these same responses can be the result of being treated as a member of a particular group when this is not one's preference. Therefore, one goal of this research is to determine whether social categorization threat influences participants' self-evaluations.

Implicit versus explicit use of a social category

A second goal of this research is to examine whether the implicit use of an inappropriate social categorization leads to different psychological consequences in comparison to the explicit use of the same category. The affective impact of prejudice and discrimination depends on the degree to which this treatment is blatantly imposed (Barreto & Ellemers, 2005a,b; Major, Quinton, & Schmader, 2003). Because blatant instances of discrimination are likely to elicit coping responses that are directed at the source of negative treatment (e.g. anger and protest intentions), it leaves the individual self comparatively unharmed. By contrast, the ambiguity inherent in implicit forms of prejudice impairs negative other-directed responses, and instead tends to elicit negative self-directed affect (such as anxiety, or low personal self-esteem). The examination of responses to subtle versus blatant discrimination often implies keeping reference to *group membership* constant and explicit, and varying only whether *in-group inferiority* is explicitly stated or implicitly implied (e.g. Barreto & Ellemers, 2005a,b; Major *et al.*, 2003). Although the process we propose here is similar, we now examine how targets respond to group-based treatment when categorization is implicit versus explicit. We hypothesize that explicit categorization threats are likely to elicit rejection of the categorization, as reflected in anger and protest, whereas implicit categorization threats are more likely to result in a feeling of personal inadequacy (because one's identity preferences are not taken into account and one cannot pinpoint exactly why), as reflected in negative self-directed affect and low personal self-esteem.

Overview of the studies

We report three studies in which we examine the extent to which participants express rejection of implicit and explicit unwanted categorizations and the impact of this treatment on the individual self. In Studies 1 and 2, we also examine whether treatment source affects reactions to unwanted categorizations. Prior research suggests that

people tend to be more tolerant of negative treatment from in-group members in comparison to out-group members (Baron, Burgess, & Kao, 1991; Barreto & Ellemers, 2005a; Ellemers, Van Rijswijk, Bruins, & de Gilder, 1998; Huo & Tyler, 2001; Tyler & Blader, 2002). Also, people tend to be more suspicious of the motives of out-group members than of the motives of in-group members (Duck & Fielding, 2003; Schopler & Insko, 1992). Importantly, because an out-group member may make the intergroup context more salient (Turner *et al.*, 1987), it might be easier for participants in the implicit conditions to see the treatment that they receive from an out-group member as category-based (when categorization is explicit, it would be clear to all participants). In other words, participants' reactions to an implicit categorization imposed by an out-group source may come closer to their reactions to explicit categorization. In addition, an out-group source may elicit more hostile reactions than an in-group source simply because negative behaviour in an intergroup context may build on existing intergroup tensions. To explore this possibility, we manipulated the gender of the team member responsible for allocating tasks to other participants.

In Study 1, we examine the impact of contextually inappropriate categorizations on the individual self by examining negative self-directed affect. In Study 2, we examine rejection of categorization with a wide range of indicators: agreement with the task allocation, reported anger, willingness to protest, and self-stereotyping. In Study 3, we examine whether reactions to implicit versus explicit contextually inappropriate categorization are modified by whether or not categorization provides the target with an advantage. Here, we examine both rejection of categorization and the impact of the treatment on the individual self. In this study, rejection of categorization is examined with agreement with the task allocation, and reported anger, whereas the impact of the treatment on the self is examined with negative self-directed affect and personal self-esteem.

STUDY 1

In this study, we developed a paradigm which involves the allocation of proof-reading texts among group members. Participants, who were all female, first stated their task preference but subsequently were assigned stereotypically female tasks, irrespective of their preference. We varied whether or not gender was explicitly indicated as criterion for task allocation, and whether the tasks were allocated by a male or a female source. We expected that – when explicitly asked – participants would indicate that gender was an irrelevant criterion for this task allocation. To insure that female stereotypical texts were not less attractive in comparison to the male stereotypical or neutral texts, we first conducted a pilot test with a smaller sample from the same population as the main study.

Pilot Study

Because the main experimental task required participants to choose among stereotypical feminine, masculine and neutral texts to proofread, we asked a separate sample of twenty six female students to rate the stereotypicality and the attractiveness of 59 titles. Thirteen participants were asked to rate the gender stereotypicality suggested by each title (from ' − 3' *typically male* to '0' *neutral* to '3' *typically female*). Thirteen other participants rated the attractiveness of each title (from ' − 3' *totally*

unattractive to '0' *neutral* to '3' *totally attractive*). We used these ratings to select text titles that were rated as similar in attractiveness but clearly different in stereotypicality.

We selected two stereotypically female ('Diet: An Easy Loss' and 'Childcare: The Health of Your Child'), two neutral ('Career: The Possibilities in Your Career' and 'Photography: A Photographic View'), and two stereotypically male titles ('Adventure: The Ascent of Mount Everest' and 'Science Fiction: Aliens, Fact or Fiction?'). Stereotypicality ratings show that all sets of titles differed in the extent to which they were seen as stereotypically male or female (female stereotypical titles: $M = 1.73$, $SD = .59$; neutral titles: $M = .12$, $SD = .46$; male stereotypical titles: $M = -1.38$, $SD = .62$), all $t(12) > 6$, all $p < .001$. Moreover, comparisons of stereotypicality ratings to the mid-point of the scale indicate that the female stereotypical titles were rated as female stereotypical (i.e. means reliably above the mid-point of the scale, $t(12) = 10.42$, $p < .001$), the male titles were rated as male stereotypical (i.e. means reliably below the mid-point of the scale, $t(12) = -8.08$, $p < .001$), and the neutral titles were rated as neutral (i.e. means statistically equal to the mid-point of the scale, $t(12) = .89$, *ns*).

The female stereotypical titles selected were rated as equally attractive as the neutral titles (female stereotypical titles: $M = 1.12$, $SD = 1.04$; neutral titles: $M = 1.42$, $SD = .86$), $t(12) = -.69$, *ns*. However, the two male stereotypical titles selected were rated as less attractive than the neutral titles (male stereotypical titles: $M = -1.58$, $SD = 1.47$; neutral titles: $M = 1.42$, $SD = .86$), $t(12) = -3.87$, $p < .01$ and the female stereotypical titles (female stereotypical titles: $M = 1.12$, $SD = 1.04$; male stereotypical titles: $M = -1.58$, $SD = 1.47$), $t(12) = 2.28$, $p < .05$. Ratings of the attractiveness of both the female stereotypical and the neutral titles were reliably above the neutral point of the scale, both $t(12) > 3.85$, $p < .01$, showing that they were seen as relatively attractive. However, ratings of the attractiveness of the male titles were not reliably different from the neutral point of the scale $t(12) = -.34$, *ns.* Still, the lower attractiveness ratings for the male stereotypical titles should – if anything – make the assignment of female stereotypical titles more appealing and perhaps increase participants' acceptance of the category based treatment, and thus implies a conservative test of our hypothesis.

Main Study

Method

Design and participants

The design was a 2 (categorization: implicit vs. explicit) × 2 (gender of source: female vs. male) between participants factorial design. Sixty female students of Sonoma State University (a public state University in California, USA) took part in this experiment. The experiment lasted approximately 20 min, after which participants were fully debriefed. All participants had a chance to win 100 USD in a lottery. A few participants who were students in introductory psychology classes received credit for participating.

Procedure

Participants sat in front of a computer through which all instructions were provided. Participants read that they would be part of a work team with the task of correcting texts on spelling and grammar. To preserve anonymity, all participants would be allocated a false name that corresponded to their gender. Participants first indicated their gender,

and were subsequently allocated the false name 'Maria'. The next computer screen showed two male work team members (Pete & John) and one female team member (Jennifer). In reality all information about other team members was pre-programmed and all participants in the study were female.

Next, participants viewed a list with the titles of six texts that had to be corrected. The texts were presented in random order. Participants read that they would be asked to correct one of these texts, and that one team member would be randomly chosen by the computer to allocate these texts among all team members. This constituted the manipulation of gender of source: in the *female source* condition, the computer subsequently selected Jennifer to allocate the texts, while in the *male source* condition, John was selected. Next, participants had the opportunity to indicate which texts they wished to work on – they could indicate their first and second preferences. Participants read that the team member allocating the tasks (John or Jennifer, depending on experimental condition) could take these preferences into account when deciding how to allocate the texts among team members. Other than this, participants remained unaware of what other information or instructions had been provided to the task allocator.

Participants then saw on the screen which text had been allocated to the team members. The names of all team members appeared on the left side of the screen, and on the right side of the screen the title allocated to each member appeared. Participants could see that Pete had been assigned the text 'Adventure: The Ascent of Mount Everest', John had been assigned 'Science Fiction: Aliens, Fact or Fiction?', Jennifer had been assigned 'Diet: An Easy Loss', and the participant herself had been assigned 'Childcare: The Health of Your Child'. We chose to allocate this particular title to the participant because many students at this university are interested in child care and developmental psychology (many psychology students baby-sit on a regular basis, work in day-care part-time, and some are even mothers themselves). Of the two possible titles, we felt this title would be less likely to elicit negative self-direct affect due to insecurity about lack of knowledge or ability.

Participants were asked to write down on their task form which text they were to correct. In the *explicit categorization* condition, the next screen indicated the source's motivation for this task allocation: 'John/Jennifer gives you this text because he/she thinks that it is appropriate for you, because you are a woman'. In the *implicit categorization* condition the (bogus) source's motivation was: 'John/Jennifer gives you this text because he/she thinks that it is appropriate for you, because of who you are'. We reasoned that this phrase would suggest that the allocation decision reflected a participant characteristic (possibly including gender) but keep the exact criterion implicit.

Dependent measures

Two manipulation checks were asked just after the manipulation: participants indicated the gender of the team member that had allocated the texts (check of the manipulation of gender of source), and which text had been assigned to them (check of the manipulation of categorization). At the end of the experiment, we also checked that participants agreed that gender was an irrelevant criterion for this task distribution (relevance of gender as task distribution criterion). Responses were indicated on 7 point Likert scales (from '1' *not at all* to '7' *very much*). *Negative self-directed affect* was assessed with the following emotions: self-confident, self-assured, and strong (responses were indicated on 7 point Likert scales, from '1' *not at all* to '7' *very much*; all items were recoded so that higher scores indicate more negative affect, $\alpha = .63$).

Results

Manipulation checks

All participants indicated correctly which text they had been allocated (this was true in all studies). Only one participant (in the male source/explicit categorization condition) indicated the wrong source. Similar results are obtained when the data are analyzed with and without this participant, so the data for this participant were retained. Importantly, across all conditions, participants did not think that gender was a relevant criterion for task distribution (overall $M = 2.52$, $SD = 1.71$). No effects of the manipulations were obtained on this measure, indicating that all participants thought that gender was irrelevant for task distribution in this context.

Covariate: Matching between self-chosen and external categorization

To correct for any differences in initial task preference, we computed a score reflecting the extent to which participants chose stereotypical feminine titles (i.e. the extent to which self-chosen and external categorizations matched).[2] Participant's task preferences were coded as '0' if they chose two non-stereotypical feminine titles, '1' if they chose one stereotypical feminine title and '2' if they chose two stereotypical feminine titles. In total, 42% of the participants did not choose any female stereotypical title, 38% of the participants chose one female stereotypical title, and 20% of the participants chose two female stereotypical titles.

Negative self-directed affect

A 2 (categorization) × 2 (gender of source) ANCOVA revealed only a reliable Categorization effect, $F(1, 55) = 6.14$, $p < .05$. As predicted, participants who were categorized in an implicit manner ($M = 3.66$, $SD = 1.10$) reported more negative self-directed affect than participants who were explicitly categorized ($M = 2.93$, $SD = .93$, see Table 1).

Discussion

The results of this first study offer an initial demonstration of how implicit categorization threats can harm the individual self. Even though all participants reported that gender was *not* a relevant criterion for task distribution, implicitly categorized participants reported more self-directed affect than participants who were explicitly categorized. In contrast to previous research on stereotype threat (see Schmader *et al.*, 2008, for a recent review), being categorized as a woman in this context should not have been associated with lower expectations for the experimental task – if anything, participants could have expected to perform better on the female stereotypical topics (with which they could have more experience and pilot participants rated as more attractive) in comparison to the male or neutral topics. Nevertheless, being implicitly categorized as women was a negative experience in this context. In other words, being treated on the

[2] *We also created a second co-variate in which we coded whether participants chose the title that was assigned to them (task-match). We obtain the same results as reported in this paper with one exception: the result for self-stereotyping in Study 2 is entirely reliable if we use the task-match score, but only marginally reliable if we use the stereotype-match score.*

Table 1. Means for main effects of categorization (all studies)

	Categorization	
	Implicit	Explicit
Resistance to the categorization		
Agreement (Study 2)	3.86 (1.63)[a]	2.84 (1.68)[b]
Agreement (Study 3)	4.49 (1.74)[a]	3.25 (1.46)[b]
Anger (Study 2)	2.72 (1.70)[b]	3.78 (1.75)[a]
Anger (Study 3)	2.38 (1.42)[b]	3.44 (2.02)[a]
Protest (Study 2)	3.64 (1.09)[b]	4.39 (1.27)[a]
Self-views		
Negative self-directed affect (Study 1)	3.66 (1.00)[a]	2.93 (0.91)[b]
Negative self-directed affect (Study 3)	2.92 (1.01)[a]	2.40 (1.01)[b]
Personal self-esteem (Study 3)	4.69 (1.04)[b]	5.27 (1.09)[a]

Note. Scores range from 1 to 7. Only means with different superscripts differ from each other ($p < .05$). Comparisons are made within rows. Standard deviations are presented in parentheses.

basis of a contextually inappropriate category is problematic even when it may be associated with positive expectations.

In this study, we did not find any effect of group membership of the source of treatment. It may be that group membership of the source did not affect categorization salience. Alternatively, the source manipulation may have affected categorization salience, but categorization salience did not shape the influence of implicit or explicit categorization threat on negative self-directed affect. However, treatment source group membership may affect other types of responses, such as anger towards the source of treatment. Study 2 was conducted to extend these results and includes an examination of whether or not treatment source affects categorization salience as well as reported anger.

STUDY 2

We designed this study with three objectives. First, we used direct indicators of categorization rejection including the extent to which participants agree with the task allocation, their reported anger and their willingness to protest the task allocation. In contrast to the first study in which we asked participants to indicate whether *gender* should be relevant for task allocation, in this study participants indicated whether they agree with the task allocation, without any reference to gender. These two measures are quite different. The measure included in Study 1 allows us to conclude that, when explicitly asked, participants see gender as an irrelevant task allocation criterion. However, the measure included in this second study is an indicator of people's ability to express disagreement with the way they are treated, and this should depend on whether the categorization is implicit or explicit. That is, whether or not participants make the connection between the way they are treated and their gender is likely to depend on the explicitness of the categorization. This means that in this study we expect participants to agree more with the task allocation when it is implicit than when it is explicit. We also include anger and protest among our indicators of identity resistance. This allows us to examine a possible effect of group membership of the treatment source, given that

out-group sources may elicit greater hostility than in-group sources, because they build on existing intergroup tensions.

Second, we examined the impact of contextually inappropriate categorizations on self-descriptions, and specifically on self-stereotyping. Research on self-verification versus behavioural confirmation processes shows that people reject external views of themselves with which they disagree when this discrepancy is clear, but fail to do so when the discrepancy is unclear (Swann & Ely, 1984; see also Kray, Thompson, & Galinsky, 2001). In addition, when the discrepancy is unclear, people may actually display behaviour that unwittingly confirms the external self-view. As a consequence, and applying this reasoning to the context we are examining, people may be unable to reject inappropriate categorizations when these are implicitly imposed and even respond in ways that unwittingly confirm the appropriateness of that categorization. We investigate this process by examining the extent to which participants describe themselves in stereotype consistent ways after they have been categorized implicitly or explicitly. Our hypothesis is that when targets are explicitly exposed to an inappropriate categorization they are more able to reject this categorization and avoid describing themselves in stereotype consistent ways, compared to when the categorization is implicit.

Finally, in this study we introduce some methodological variations to further test the robustness of our findings. In Study 1, the treatment source indicated that there was something about the target which justified the task allocation. However, implicit categorizations are not always accompanied by justifications. Although we see benefits in the procedure we used in Study 1, it may also be that this justification directed the participant's attention to her individual shortcomings, explaining the negative self-directed affect and possibly reducing resistance to the categorization. To examine whether implicit categorizations that are unaccompanied by a justification have similar effects, Study 2 (and Study 3) do not include any motivation for the task allocation in the implicit condition, whereas the motivation in the explicit condition remains the same as in Study 1. In addition, the manipulation checks used in Study 1 may have cued the study's goals (especially the gender source manipulation check), so they are now included only at the end of the study. We also check the categorization manipulation in two different ways: by asking to what extent the task allocation had anything to do with gender, and by requesting participants to make attributions to the source's worldview for the task allocation. Since attributions to group membership include seeing the source as holding prejudiced beliefs, we should find that those submitted to an explicit categorization are more likely to attribute the treatment they receive to the source's worldview than participants who are categorized implicitly. To make gender a more obvious reason for the task allocation, participants were assigned two titles to proofread (both stereotypically feminine), instead of only one title (as was the case in Study 1). For this reason we also offered a choice between 12 titles, instead of between only 6 titles.

Pilot Study

Because this study was conducted in a different country and language from the first study, we conducted a second pilot study (in exactly the same way as the first pilot study). Fifteen female participants from Leiden University rated 59 titles for stereotypicality (from '−3' *typically male* to '0' *neutral* to '3' *typically female*), and fifteen participants rated the same titles for attractiveness (from '−3' *very unattractive*

to '0' *neutral*, and to '3' *very attractive*). The stereotypically female titles chosen as stimuli for the main study were: 'Cooking: Good cooking for everyday', 'Childcare: The Health of Your Child', 'Relationships: Breaking up', and 'Spirituality: Spiritual thinking and feeling'. The male titles were: 'Adventure: The Ascent of Mount Everest', 'Aliens: Fact or Fiction?', 'Do it yourself: Build your own house', and 'Humour: The best jokes for parties'. The neutral titles were: 'Career: The Possibilities in Your Career', 'Photography: A Photographic View', 'Biography: The life of Nelson Mandela', and 'Food and drinks: Which wine with what dish?'.

Ratings of stereotypicality show that all sets of titles differed in the extent to which they were seen as stereotypically male or female (female stereotypical titles: $M = 1.45$, $SD = .75$; neutral titles: $M = -.20$, $SD = .67$; male stereotypical titles: $M = -1.23$, $SD = .68$), all $t(14) > 5$, all $p < .001$. Moreover, comparisons of stereotypicality ratings to the mid-point of the scale indicate that the female stereotypical titles were rated as female stereotypical (i.e. means reliably above the mid-point of the scale, $t(14) = 7.54$, $p < .001$), the male titles were rated as male stereotypical (i.e, means reliably below the mid-point of the scale, $t(14) = -6.98$, $p < .001$), and the neutral titles were rated as neutral (i.e. means statistically equal to the mid-point of the scale, $t(14) = -1.16$, *ns*).

The female stereotypical titles selected were rated as equally attractive as the neutral titles (female stereotypical titles: $M = .88$, $SD = .99$; neutral titles: $M = .52$, $SD = .87$), $t(14) = 1.25$, *ns*. The male stereotypical titles selected were also rated as equally attractive as the neutral titles (male stereotypical titles: $M = .38$, $SD = .84$; neutral titles: $M = .52$, $SD = .87$), $t(14) = .42$, *ns.* The female stereotypical titles were rated as marginally more attractive than the male titles (female stereotypical titles: $M = .88$, $SD = .99$; male stereotypical titles: $M = .38$, $SD = .84$), $t(14) = 1.87$, $p = .08$. Also, ratings of the attractiveness of both the female stereotypical and the neutral titles were reliably above the neutral point of the scale, both $t(14) > 2.3$, $p < .05$, showing that they were seen as relatively attractive. However, ratings of the attractiveness of the male titles were only marginally above the neutral point of the scale $t(14) = 1.77$, $p = .09$. Note, however, that the slightly lower attractiveness of the male stereotypical titles should make participants more accepting of the category based treatment, and thus implies a conservative test of our hypotheses.

Main Study

Method

Design and participants

The design was again a 2 (categorization: implicit vs. explicit) × 2 (gender of source: female vs. male) between participants factorial design. Eighty eight female students from Leiden University, The Netherlands, participated in the experiment (approximately 22 per condition) and were paid the equivalent of 1 USD for participating.

Procedure

The procedure of this study was the same as in Study 1 except for two aspects. First, in this study, participants were allocated two texts instead of just one (so they saw a list of 12 titles). The texts allocated to the participant in this study were on the topics of

child care and cooking. Again, both of these topics are topics with which our student population is familiar. Second, in this study no motivation was provided for the task allocation made by the source in the implicit condition, while in the explicit condition the same motivation was given as in Study 1 (i.e. 'Henk/Anne gives you this text because he/she thinks that it is appropriate for you, because you are a woman').

Dependent measures

Unless otherwise indicated, all questions consisted of statements with which participants agreed or disagreed (from 1 '*not at all*' to 7 '*very much*'). To check the categorization manipulation first we measured *attributions of the task allocation to the source's worldview* with two items: to the worldview of the team member who was assigned to allocate the texts and to the personality of the team member who was selected to allocate the texts ($r = .82$, $p < .01$). We also asked two direct manipulation questions at the end of the experiment (when cueing gender was less problematic). To check the categorization manipulation, participants indicated whether they thought that the task allocation had anything to do with the fact that they were woman. To check the gender source manipulation, participants indicated the gender of the team member who had allocated the tasks (male vs. female).

Agreement with the task allocation was assessed with two items inquiring about the extent to which participants agreed with the task allocation and the extent to which it was fair ($r = .73$, $p < .001$). *Anger* was measured with four items: angry, irritated, indignant, and hostile ($\alpha = .89$). *Willingness to protest* was measured with five items (e.g. If we have to do another task, I would like the same team member to allocate the tasks (recoded), I wish to speak to the experimenter; $\alpha = .72$). *Self-stereotyping* was assessed with a scale that we developed for this study. This scale was based on the Stereotype Avoidance Scale by Steele and Aronson (1995). The scale included 3 activities that are seen as stereotypically associated with the stereotype of women (to take dance lessons, to go shopping, and to apply make-up, $\alpha = .73$).[3] Participants were asked to indicate the extent to which they enjoy each of these activities. The selection of these activities was based on additional piloting with fifteen female participants from the same population as the main study. The results of the pilot study confirm that these activities were seen as stereotypically female ($M = 2.00$, $SD = .67$ on a scale from -3 = *stereotypically male* to $+3$ = *stereotypically female*), and that they were regarded as attractive ($M = 1.78$, $SD = .64$ on a scale ranging from -3 = *very unattractive* to $+3$ = *very attractive*).

Results

Manipulation checks

Five participants indicated the gender of the source incorrectly. These five individuals were in the female source conditions. Similar results are obtained with and without these participants so these data were retained.

A 2 (categorization) × 2 (gender of source) analysis of variance (ANOVA) on attributions for the task allocation revealed a main effect of categorization,

[3] *We also piloted and included 3 stereotypically male activities, and 3 stereotypically neutral activities. However, these scales were unreliable.*

$F(1, 84) = 7.99$, $p < .01$. Participants in the explicit condition attributed the way the texts were allocated more to the source's worldview ($M = 5.44$, $SD = 1.81$) than participants in the implicit condition whose score around the scale mid-point indicated they did not have a strong opinion about whether or not this was the case ($M = 4.39$, $SD = 1.68$). The extent to which participants thought that their gender had something to do with the task allocation was also affected by Categorization, $F(1, 83) = 14.81$, $p < .001$. Participants who had been explicitly categorized more strongly endorsed the idea that their gender had something to do with the allocation of the texts ($M = 5.93$, $SD = 1.56$) than participants who were implicitly categorized who again did not have a strong opinion about whether or not this was the case ($M = 4.36$, $SD = 2.22$). These results indicate that our manipulation of categorization was successful. Note also that the means for the implicit condition are around the scale mid-point (both $t(43) < 1$, *ns*), denoting the uncertainty participants in this condition feel about what caused the treatment they received.

Covariate: Matching between self-chosen and external categorization

The covariate was computed in the same way as in Study 1. In this study, 36% of participants chose two titles that were not stereotypically female (only stereotypically male or neutral), 39% chose one stereotypically female title, whereas 24% chose two stereotypically female titles.

Agreement with the task allocation

An ANCOVA showed a reliable main effect of Categorization, $F(1, 83) = 6.89$, $p < .01$, indicating that participants in the implicit condition agreed more with the (gender-based) task allocation ($M = 3.86$, $SD = 1.63$) than did participants in the explicit condition ($M = 2.84$, $SD = 1.68$). Participants in the implicit condition again scored around the scale mid-point, suggesting they did not have a strong opinion about whether or not they disagreed with the task allocation.

Anger

Participants who had been explicitly categorized reported more anger ($M = 3.78$, $SD = 1.75$) in comparison to participants categorized in an implicit way ($M = 2.72$, $SD = 1.69$), $F(1, 83) = 8.66$, $p < .005$. Also, irrespective of categorization, a male source generally elicited more anger ($M = 3.74$, $SD = 1.75$) in comparison to a female source ($M = 2.72$, $SD = 1.69$), $F(1, 83) = 7.20$, $p < .01$. Note however that this effect did not qualify the main effect of categorization. Note also that although the scores on this measure are rather low, scores on negative emotions in social psychological research are typically quite low, and should thus be seen more in comparison to each other than in absolute terms.

Willingness to protest

A 2 (categorization) × 2 (gender of source) analysis of covariance (ANCOVA) revealed a reliable main effect of categorization, $F(1, 83) = 8.08$, $p < .01$. Participants who had been explicitly categorized had a stronger wish to protest ($M = 4.45$, $SD = 1.24$) in comparison to participants categorized in an implicit way ($M = 3.75$, $SD = .99$). Again, we also obtained a reliable main effect of Source, $F(1, 83) = 5.55$, $p < .05$. A male

source elicited a stronger wish to protest ($M = 4.39$, $SD = 1.13$) in comparison to a female source ($M = 3.81$, $SD = 1.15$). Note however that this effect did not qualify the main effect of categorization.[4]

Self-stereotyping

An ANCOVA revealed a marginally reliable interaction between categorization and gender of source on self-stereotyping, $F(1, 83) = 3.58$, $p = .06$. Simple effects show that the predicted effect of categorization is reliable when the source is male, but not when the source is female (see Table 2). As a consequence, and consistent with our predictions, participants' self-stereotyped to a greater extent when the categorization was implicit than when it was explicit. However, this only happened when the source was male, and not when it was female. This finding is likely to be due to the greater salience of gender identity that is elicited by a context with a male source. By contrast, a female source renders gender identity less salient, making participants less vulnerable to self-stereotyping in the implicit conditions.

Discussion

This study extended the results of Study 1 and our understanding of the effects of contextually inappropriate categorizations, as a function of whether they are implicit or explicit. First, we tested the effects of our experimental paradigm more extensively and found results indicating the success of our manipulations. Noteworthy are the results on our measure of attribution for the task allocation, which demonstrate that participants made more external attributions when treatment was due to explicit categorization than when this was implicit.

We again demonstrated that targets of inappropriate categorizations are less likely to reject this treatment when it is implicit. Indeed, participants expressed more disagreement, more anger and greater wish to protest when the inappropriate categorization was explicit in comparison to implicit. Moreover, consistent with prior research within self-verification theory (Swann & Ely, 1984; see also Kray *et al.*, 2001), we found that participants in the explicit categorization condition were able to reject the imposed categorization and distance themselves from the female stereotype. However, this only happened when participants were categorized by a male source, and not when participants were implicitly categorized by a female source. A close look at the pattern of results reveals that in the implicit conditions participants self-stereotyped less when the source was female than when the source was male. This indicates that a male source might increase the salience of the gender categorization when the treatment is implicit (consistent with self-categorization theory, Turner *et al.*, 1987). If this is correct, then our findings suggest that, when categorizations are implicit, self-descriptions tend to be in line with categorizations that might be made salient by contextual cues (such as

[4] *Both agreement with the task allocation and anger at the task allocation reliably and independently mediated the effect of categorization on willingness to protest. Following the steps outlined by Baron and Kenny (1986), categorization reliably affected willingness to protest ($\beta = .31$, $p < .005$), categorization reliably affected agreement with the task allocation ($r = -.29$, $p < .005$) and when agreement was entered as a predictor at the same time as the manipulation, the effect of the manipulation dropped to non-significance ($\beta = .12$, ns), Sobel $z = 2.71$, $p < .01$. Also, categorization reliably affected anger ($r = .29$, $p < .005$) and when anger was entered as a predictor at the same time as the manipulation, the effect of the manipulation dropped to non-significance ($\beta = .08$, ns), Sobel $z = 2.79$, $p < .005$.*

Table 2. Self-stereotyping as a function of categorization and source (Study 2)

	Categorization	
	Implicit	Explicit
Gender of source		
Female	4.48[b]	4.92[b]
	(1.42)	(1.57)
Male	5.58[a]	4.76[b]
	(1.12)	(1.55)

Note. Scores range from 1 to 7. Only means with different superscripts differ from each other ($p < .05$). Standard deviations are presented in parentheses.

the group membership of the source of categorization), while these cues have a weaker influence when categorizations are explicit.

Male sources also generally elicited more anger and more protest than female sources. However, it is important to note that – with the exception of self-stereotyping – treatment source did not qualify responses to the categorization threat.

STUDY 3

The third study was conducted to determine whether the effects of categorization threat remain when this treatment offers an advantage. As we argue in the introduction, categorization threats can be associated with *positive* outcomes as well as with *negative* outcomes. Contrary to what happens with benevolent sexism, what is negative about categorization threat is not the *outcome*, but the *process* itself – the fact that one's choice of identity is not taken into account. At a theoretical level, categorization threats emerge due to a discrepancy between contextually chosen self-views and external treatment, even when this treatment is associated with an advantage. At an empirical level, evidence from past research indirectly suggests this may be the case – for example, by showing that people attach great importance to being treated according to their choice, even when they would otherwise be treated as members of a higher status group (Barreto & Ellemers, 2002). This research suggests that the possibility of receiving an advantage should not change targets' readiness to resist a social categorization that is contextually inappropriate. Similarly, research on reactions to affirmative action has shown that people who receive a positive outcome (e.g. job selection) on the basis of their social category membership rather than on the basis of personal merit often resist this treatment and suffer low self-confidence as a result (e.g. Heilman, Battle, Keller, & Lee, 1998; Heilman, Simon, & Repper, 1987). Our focus is slightly different: we do not contrast category-based treatment with merit, but instead we contrast category-based treatment with personal preference. However, we also do not expect category-based treatment to be accepted when it is prioritized over personal preference. Finally, the results from the pilot studies suggest that participants are likely to reject this unwanted categorical treatment even if it offers an advantage. Participants in the pilot studies rated the stereotypically male tasks as relatively less attractive than the stereotypically female tasks, but study participants still resisted being asked to perform female stereotypical tasks. However, past research also shows that favourable outcomes often are judged as

fair (e.g. Van den Bos, Wilke, Lind, & Vermunt, 1998), even when they are associated with category-based treatment (Ellemers & Barreto, 2006). Therefore, in the third study, we directly examine whether participants will resist category-based task allocations, even if the task offers them an advantage, by experimentally manipulating whether or not the female stereotypical tasks allocated to participants are more attractive in comparison to the remaining tasks.

In this study, we examined both categorization rejection and the impact of the unwanted treatment on the self. Categorization rejection was examined as agreement with the task allocation and reported anger. The impact of the categorization on the self was examined in this study with negative self-directed affect and a measure of personal self-esteem. This adds one important indicator of the impact of categorization threats on the self to those investigated so far. In fact, personal self-esteem is a crucial indicator of an individual's self-views, and one that determines individual behaviour in a variety of contexts (see, e.g. Baumeister, 1998).

In contrast to the previous two studies, the gender of the team member who assigns the titles was not manipulated. Instead, a male team member always assigned the titles. We did this because the source of categorization did not qualify the effects of categorization in studies 1 and 2 for the measures we focus on in this study (i.e. gender of source only qualified self-stereotyping). Our focus on a male source stems from the consideration that given that male sources enhance the salience of gender identity among female participants, it could also be easier to reject the inappropriate treatment from a male than from a female source. Therefore, this constitutes a conservative test of our argument that implicit categorization threat is more difficult to reject and more damaging for the self than an explicit threat.

Method

Design and participants

The design of this study was a 2 (categorization: implicit vs. explicit) × 2 (outcome of the categorization: advantage vs. no advantage) between participants factorial design. Seventy-five female students from Leiden University participated in the experiment and were paid 1 USD for participating.

Procedure

The procedure employed in this study was similar to that of Study 2, except for two changes. First, in all conditions, the source of gender categorization was a male group member (Henk). Second, we manipulated the attractiveness of the stereotypical feminine titles allocated to participants (i.e. the Outcome manipulation). To do so, we went back to the results of our pilot test of attractiveness and stereotypicality of a set of 59 titles. From these, we selected four titles that were rated as stereotypically female (relatively to the neutral point of the scale, and relatively to the male stereotypical and the neutral titles), that could be divided in two sets of two titles each, which differed in attractiveness but not in relative stereotypicality. That is, the two titles chosen for the advantage condition should be seen as more attractive than the two titles chosen for the no advantage condition, as well as more attractive than the male stereotypical and the neutral titles. However, the female titles chosen for the two conditions should be seen as equally stereotypically female.

Following these criteria, for the advantage condition, we selected two female stereotypical titles that had been rated in the pilot study as relatively attractive

compared to the neutral point of the scale ($M = 1.40$, $SD = .89$), $t(14) = 6.09$, $p < .001$. These titles were also rated by our pilot participants as more attractive than the neutral ($M = .52$, $SD = .87$) and the male stereotypical titles ($M = .38$, $SD = .84$), both $t(14) > 3.35$, both $p < .005$. These titles were: 'Shops: the nicest shops in The Netherlands' and 'Cooking: Nice cooking for everyday'.

For the no advantage condition we selected two female stereotypical titles that had been rated by our pilot participants as neutral in attractiveness, or equal to the mid-point of the scale ($M = -.14$, $SD = .89$), $t(13) = .56$, *ns*. These titles were rated as slightly less attractive than the male stereotypical titles, $t(13) = 1.94$, $p = .07$, as well as than the neutral titles, $t(13) = 2.10$, $p = .06$. Note however that the fact that the female titles in the no advantage condition were rated as slightly less attractive than the remaining titles works against our expectation that this manipulation will not affect the extent to which participants reject the category-based treatment, and thus provides a conservative test of our hypotheses. The titles in the no advantage condition were: 'Marriage: Marriages and everything that can go wrong' and 'Ponies: taking care of your pony'. It is important to note that the female stereotypical titles selected for the advantaged condition were rated as more attractive than the female titles selected for the no advantage condition, $t(14) = 4.39$, $p < .001$. Again, all of the topics covered by the texts allocated to participants in all conditions were familiar to our student population.

The titles selected for the advantage and for the no advantage condition were rated as equally stereotypically female (advantage: $M = 1.83$, $SD = .82$; no advantage: $M = 1.77$, $SD = .92$), $t(14) = .27$, *ns.* In addition, these titles were rated as female stereotypical both compared to the neutral point of the scale, both $t(14) > 7.00$, $p < .001$, and compared to the male and the neutral titles, all $t(14) > 5.60$, all $p < .001$.

Dependent measures

Before task allocation, participants saw the list of titles that would be allocated to the team members. Then participants indicated on seven point Likert scales to what extent they found each of the titles attractive. The attractiveness of the female stereotypical texts that were later allocated to participants relative to the remaining titles (stereotypical male and neutral) was the check of the manipulation of Outcome. Two other manipulation checks were included at the end of the experiment. Participants indicated the gender of the source, as well as whether the task allocation had anything to do with the fact that they were a woman (check of the manipulation of Categorization).

Agreement with the task allocation ($r = .77$, $p < .001$) was assessed in the same way as in Study 2 and *negative self-directed affect* ($\alpha = .73$) was assessed in the same way as in Study 1. *Anger* was assessed in the same way as in Study 2 ($\alpha = .90$). *Personal self-esteem* was measured with five questions adapted from the state self-esteem scale of Heatherton and Polivy (1991). Three of these items stem from the social self-esteem subscale, reflecting our interest in how people feel about themselves as a result of how they are treated by others (at this moment, I am worried about what other people think of me, I feel inferior to others at this moment, at this moment, I am concerned about the impression I am making, all items recoded). To this we added two items, adapted from the appearance subscale, so as to assess personal self-esteem more globally (I feel pleased with myself at this moment and I am not very pleased with myself at this moment, recoded). Since these two subscales had similar patterns and formed a reliable scale together ($\alpha = .78$), we averaged scores across all items. The correlation between

the personal self-esteem and the negative self-directed affect measures was negative and reliable ($r = -.26$, $p < .05$).

Results

Manipulation checks

All participants indicated that the group member allocating the texts was male. A 2 (categorization: implicit vs. explicit) × 2 (outcome: advantage vs. no advantage) ANOVA revealed only a reliable main effect of Categorization on the manipulation check of type of categorization, showing that participants in the explicit condition stated that their gender had more to do with the task allocation ($M = 6.19$, $SD = 1.17$) than did participants in the implicit condition ($M = 3.90$, $SD = 2.02$), $F(1, 70) = 34.98$, $p < .001$. We also tested this difference with a t test, not assuming equality of variances (Levene's test, $F = 22.29$, $p < .001$), $t(61.62) = 6.08$, $p < .001$. Again, as in Study 2, the means for the implicit condition are around the scale mid-point ($t(38) = .32$, *ns*), reflecting uncertainty about the reasons underlying task allocation in this condition (which is also suggested by the greater variance in the implicit condition).

To check whether categorization outcome manipulation was successful, we compared how participants evaluated the titles that were later allocated to them with how they evaluated the remaining titles in the advantage and the no advantage conditions. As intended, in the advantage conditions, participants evaluated the titles that were later allocated to them ($M = 4.83$, $SD = 1.33$) more positively than the remaining titles ($M = 3.81$, $SD = .56$), $t(35) = 3.72$, $p < .001$. By contrast, this difference was not reliable in the no advantage condition (allocated titles: $M = 3.95$, $SD = 1.37$; remaining titles: $M = 4.04$; $SD = .87$), $t(38) = -.45$, *ns*. In addition, an ANOVA with categorization (implicit vs. explicit) and Outcome (advantage vs. no advantage) as between participants factors revealed only a reliable main effect of Outcome such that participants evaluated the titles allocated to them more positively in the advantage condition ($M = 4.86$, $SD = 1.33$, significantly above mid-point, $t(35) = 3.89$, $p < .001$), than in the no advantage condition ($M = 3.95$, $SD = 1.33$, equal to mid-point, $t(38) = .23$, *ns*).

Covariate: Match between self-chosen and external categorization

We computed a covariate in the same way as in the prior two studies. In this study, 24% of participants did not choose a stereotypically feminine title, 53% of participants chose one stereotypically feminine title and 23% of participants chose two stereotypically feminine titles.

Resistance to the categorization

Agreement with the task allocation

An ANCOVA showed a reliable main effect of categorization, $F(1, 70) = 11.67$, $p < .001$. Participants who had been implicitly categorized agreed more with the task allocation ($M = 4.49$, $SD = 1.73$) than participants explicitly categorized ($M = 3.25$, $SD = 1.46$). This result was obtained regardless of whether or not the gender-based task allocation resulted in an advantage for the participant.

Anger

Consistent with our predictions, participants who were explicitly categorized experienced more anger ($M = 3.44$, $SD = 2.02$) than those we were categorized in an implicit way ($M = 2.38$, $SD = 1.42$), $F(1, 70) = 6.48$, $p < .05$. Again, this result was independent of whether or not the gender-based task allocation resulted in an advantage for the participant.

Impact on the individual self

Negative self-directed affect

In line with our predictions, participants who were categorized in an implicit way experienced *more* negative self-directed affect ($M = 2.92$, $SD = 1.01$) than participants who were explicitly categorized ($M = 2.39$, $SD = 1.01$), $F(1, 70) = 4.78$, $p < .05$. Again, this result was independent of whether or not the gender-based categorization resulted in advantage.

Personal self-esteem

A 2×2 ANCOVA revealed a reliable main effect of categorization, $F(1, 70) = 4.27$, $p < .05$. Participants categorized in an implicit way reported *lower* personal self-esteem ($M = 4.65$, $SD = 1.17$) than participants explicitly categorized ($M = 5.23$, $SD = 1.16$). Once again, this result was independent of whether or not the gender based categorization resulted in advantage.

Discussion

The results of this study replicate and extend the results of studies 1 and 2. Participants exposed to an inappropriate categorization rejected this categorization to a greater extent (by expressing disagreement and more anger) when it was explicitly imposed than when it was imposed in an implicit way. In addition, we again found that participants exposed to an implicit categorization threat expressed more negative self-directed affect in comparison to participants exposed to an explicit categorization threat. In this study, we replicated this result with another indicator of the impact of this treatment on the self: personal self-esteem. We found that inappropriate categorizations lower the self-esteem of their targets to a greater extent when they are implicit in comparison to when they are explicit. In contrast to Study 1 in which we obtained this effect with a manipulation of categorization that always explicitly referred to the self as the cause of task allocation (but then either explicitly mentioned group membership or not), in this study there was no explicit reference to the self as the cause, but still implicit categorization led to negative self-directed affect and low self-esteem.

In the third study, the experimental design allowed us to demonstrate that the effects of categorization threat are obtained even when the female stereotypical task assignment was more attractive. Although our manipulation appears to have worked as intended, in that participants in the advantage condition found the task they had received more attractive than the remaining tasks, as predicted we did not find this factor to moderate resistance to the inappropriate categorization, nor its negative impact on the self. That is, rejection of categorization and negative feelings about the self were obtained irrespective of whether or not the gender-based categorization would

have resulted in advantage for the participant. This result seems to indicate that having one's choice of identity taken into account is more important than the benefits one can derive from alternative categorizations, at least insofar as attractiveness of tasks is concerned. It is important to note that whether category-based treatment leads to advantage or disadvantage is likely to matter in many ways, but our results show that this does not modify the experience of categorization threat. This result also further underlines the difference between reactions to categorization and reactions to discrimination: whereas prior research found that people willingly accept discriminatory treatment when it is presented in a positive tone (Barreto & Ellemers, 2005b; Moya, Glick, Exposito, & Casado, 2007), even if it leads to negative outcomes (e.g. Abrams *et al.*, 2003; Dardenne *et al.*, 2007; Glick *et al.*, 2002), the studies in this paper show that reactions to categorization threat do not seem to be moderated by the potential advantages one can obtain through this treatment.

Although we think that the setting we studied has clear parallels to real-life settings where tasks are assigned among team members, we acknowledge that its generalizability may be limited. Further research should examine whether or not these effects are parallel to those obtained when greater advantages are involved, such as access to education or to particular jobs, against which these identity threats can often be measured in real life. In addition, it would be important to examine whether similar results would be obtained when the advantages to be gained are more long-term, such as higher salary or a promotion. However, it is important to note that in these real-life-situations it is not only the advantage that can be more significant, but the same can be the case for the identity threat, which may be felt across more contexts, it may be more chronically present, and it may be expressed more visibly than in the laboratory context we studied (see, e.g. Truax, Cordova, Wood, Wright, & Crosby, 1998). We thus believe that our results offer evidence that is relevant to the understanding of these processes in a variety of real life contexts.

GENERAL DISCUSSION

Dominant meritocratic and individualistic ideologies dictate that category-based treatment is seldom accepted in many contemporary societies, rendering explicit treatment of this kind a rare occurrence. Still, categorizing people into social groups is a fairly automatic and often even socially pragmatic process (e.g. Fiske & Taylor, 1991). As a consequence, group-based perceptions are still prevalent, albeit taking a great deal more implicit forms (e.g. Dovidio, 2001; Swim, Aikin, Hall, & Hunter, 1995). The results of the three studies we report in this paper demonstrate the undermining effects that these implicit categorizations are likely to have for the self, and highlight the need to consider the specific ways in which category based treatment affects its targets in modern societies.

Specifically, our results show that targets exposed to implicit categorization threats show less resistance to this categorization than targets exposed to explicit categorization threats. We demonstrated this effect on a range of indicators, such as agreement with the categorization, anger, and willingness to protest. In addition, our results show that implicit categorization threat is associated with more negative self-directed affect and lower personal self-esteem than when this threat is explicit. This was demonstrated across three studies, with samples from two culturally different populations and different stereotypical tasks (topics). These results demonstrate the

pernicious effects of categorization threats, irrespective of how explicitly they are imposed, as both types of threat led to negative emotions. However, these results also suggest that implicit categorization threats are more undermining to the target's self, as they elicit negative affect that undermines self-image (negative self-directed affect and low self-esteem; see also Barreto and Ellemers (2005a) for a similar analysis as a result of exposure to subtle versus blatant prejudice). The implications of these results are even clearer if we are reminded of the important role self-directed affect (such as self-esteem) has as a precursor for coping and performance (Brockner, 1988; Lazarus, 1966).

Our results also show that implicit categorization threats can actually lead to behaviour or self-definition in line with the inappropriate categorization. Although targets of explicit categorization resisted defining themselves in line with the group's stereotype, targets of implicit categorization self-defined in line with the group's stereotype when its salience was cued by the social context (i.e. when the source of categorization was male). In another way, whereas self-stereotyping reflected the contextual salience of gender identity in the implicit conditions, participants were more able to strategically counter-act this psychological salience when gender categorization was made explicit. This finding is consistent with results from prior research on self-verification processes showing that people's self-image only prevails when the discrepancy between their self-image and the image others hold of them is clear (Swann & Ely, 1984; see also Kray *et al.*, 2001).

At this point it may be important to consider whether it would be beneficial to compare the effects of implicit and explicit categorization threats to responses in a control condition, to ascertain whether the ratings found in any of these conditions differ from baseline. Although this is in principle an important endeavour, we wonder how to design such a control condition. On the one hand, this could be a condition where participants are not categorized – in which case it would differ from the experimental conditions not only because participants are not categorized but also in the content of the tasks that they are asked to perform. This would perhaps not be entirely un-confounded so also not entirely conclusive. A more appropriate control condition could be a condition where participants are categorized but this is consistent with their own wishes. We do take this latter situation into account in our procedure, by controlling for the extent to which the allocated tasks match the participants' preferences. We thus think that a comparison with a control condition could be useful, but that we already provide some information in the current paper regarding this comparison.

It is important to acknowledge that, given that our focus was on women who were categorized on the basis of their gender in a context where this treatment was inappropriate, we cannot demonstrate that the effects found are generalizable to categorization in other groups. Specifically, and given that women are stereotypically expected to be more insecure and less assertive than men, it is possible that all of our findings (on resistance to categorization and on impact on the self) merely reflect assimilation to the stereotype in the implicit conditions and stereotype reactance in the explicit conditions. Although this is indeed part of the process we propose, and an interesting phenomenon in its own right, our argument is that the effects we found are connected to categorization threat, a threat that can in principle take place with other categorizations too (e.g. as a man, as an American). The evidence obtained in Study 3 does show that our findings are not restricted to categorizations associated with disadvantage. However, future research should investigate more directly whether or not the processes uncovered here are generalizable to other types of categorizations, such as when men are categorized as such in a context where gender is irrelevant.

Another point worth noting is that our examination of the role of identity of source did not reveal entirely consistent effects across measures and studies. Identity of the source did not affect negative self-directed affect in Study 1, but in Study 2 it affected anger and protest. This supports our reasoning that out-group sources may elicit greater hostility across the board, likely because reactions to negative treatment from out-group sources build on existing intergroup tensions. In addition, identity of source interacted with categorization to affect self-stereotyping in Study 2. This supports the idea that out-group sources can increase the salience of the intergroup context, when it is not already salient (as when the categorization is explicit). This however did not modify participants' responses to categorization threat. Our findings thus suggest that an implicit identity threat directs people's attention towards the self, whereas an explicit threat directs people's efforts towards rejecting this treatment, which is inappropriate irrespective of the identity of the source, although it may elicit more anger and protest if the source is an out-group member.

We believe that our results offer important insights into how people manage and negotiate social identities within modern societies. Crucially, our results show that when people can not manage their identities in ways that they prefer, whether their categorical treatment is implicit or explicit determines not only how they feel about themselves, but also their degree of resistance.

Acknowledgements

We thank Jolanda Jetten and three anonymous reviewers for their careful and helpful comments on prior versions of this paper. This research was made possible through funding from the Dutch Science Foundation (NWO, Vernieuwingsimpuls) awarded to the first author. Manuela Barreto was at Leiden University – Institute for Psychological Research (LU-IPR) when this research was conducted.

References

Abrams, D., Viki, G. T., Masser, B., & Bohner, G. (2003). Perceptions of stranger and acquaintance rape: The role of benevolent and hostile sexism in victim blame and rape proclivity. *Journal of Personality and Social Psychology*, *84*, 111–125.

Baron, R. S., Burgess, M. L., & Kao, C. F. (1991). Detecting and labeling prejudice: Do female perpetrators go undetected? *Personality and Social Psychology Bulletin*, *17*, 115–123.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Barreto, M., & Ellemers, N. (2000). You can't always do what you want: Social identity and self-presentational determinants of the choice to work for a low status group. *Personality and Social Psychology Bulletin*, *26*, 891–906.

Barreto, M., & Ellemers, N. (2002). The impact of self-identities and treatment by others on the expression of loyalty to a low status group. *Personality and Social Psychology Bulletin*, *28*, 493–503.

Barreto, M., & Ellemers, N. (2003). The effects of being categorized: The interplay between internal and external social identities. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 14, pp. 139–170). Chichester: Wiley.

Barreto, M., & Ellemers, N. (2005a). The perils of political correctness: Responses of men and women to old-fashioned and modern sexist views. *Social Psychology Quarterly*, *68*, 75–88.

Barreto, M., & Ellemers, N. (2005b). The burden of benevolent sexism: How it contributes to the maintenance of gender inequalities. *European Journal of Social Psychology*, *35*, 633–642.

Baumeister, R. F. (1998). The self. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 680–740). New York: McGraw-Hill.

Bourhis, R. Y., Moise, L. C., Perreault, S., & Senecal, S. (1997). Towards an interactive acculturation model: A social psychological approach. *International Journal of Psychology*, *32*, 369–386.

Branscombe, N., Ellemers, N., Spears, R., & Doosje, B. (1999). The context and content of social identity threat. In N. Ellemers, R. Spears, & B. Doosje (Eds.), *Social identity: Context, commitment, content*. Oxford: Blackwell.

Branscombe, N. R., Schmitt, M. T., & Harvey, R. D. (1999). Perceiving pervasive discrimination among African Americans: Implications for group identification and well-being. *Journal of Personality and Social Psychology*, *77*, 135–149.

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

Brockner, J. (1988). *Self-esteem at work: Theory, research, and practice*. Lexington, MA: Lexington Books.

Clark, R., Anderson, N., Clark, V. R., & Williams, D. R. (1999). Racism as a stressor for African Americans: A biopsychosocial model. *American Psychologist*, *54*, 805–816.

Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: Consequences for women's performance. *Journal of Personality and Social Psychology*, *93*, 764–779.

Deaux, K. (1996). Social identification. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 777–798). New York: The Guilford Press.

Dovidio, J. F. (2001). On the nature of contemporary prejudice: The third wave. *Journal of Social Issues*, *57*, 829–849.

Duck, J. M., & Fielding, K. S. (2003). Leaders and their treatment of subgroups: Implications for evaluations of the leader and the superordinate group. *European Journal of Social Psychology*, *33*, 387–401.

Ellemers, N., & Barreto, M. (2006). Categorization in everyday life: The effects of positive and negative categorizations on emotions and self-views. *European Journal of Social Psychology*, *36*, 931–942.

Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual Review of Psychology*, *53*, 161–186.

Ellemers, N., van Rijswijk, W., Bruins, J., & de Gilder, D. (1998). Group commitment as a moderator of attributional and behavioral responses to power use. *European Journal of Social Psychology*, *28*, 555–573.

Fiske, S., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Gaertner, S. L., Rust, M. C., Dovidio, J. F., Bachman, B. A., & Anastasio, P. A. (1994). The contact hypothesis: The role of a common ingroup identity on reducing intergroup bias. *Small Group Research*, *22*, 267–277.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*, 491–512.

Glick, P., Sakalli-Ugurlu, M., Ferreira, M. C., & Souza, M. A. (2002). Ambivalent sexism and attitudes toward wife abuse in Turkey and in Brazil. *Psychology of Women Quarterly*, *26*, 292–297.

Haunschild, P. R., Moreland, R. L., & Murrell, A. J. (1994). Sources of resistance to mergers between groups. *Journal of Applied Social Psychology*, *24*, 1150–1178.

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology*, *60*, 895–910.

Heilman, M. E., Battle, W. S., Keller, C. E., & Lee, R. A. (1998). Type of affirmative action policy: A determinant of reactions to sex-based preferential selection? *Journal of Applied Psychology*, *83*, 190–205.

Heilman, M. E., Simon, M. C., & Repper, D. P. (1987). Intentionally favored, unintentionally harmed? Impact of sex-based preferential selection on self-perception and self-evaluations. *Journal of Applied Psychology*, *72*, 62–68.

Hewstone, M., & Brown, R. J. (1986). Contact is not enough: An intergroup perspective on the 'contact hypothesis'. In M. Hewstone & R. J. Brown (Eds.), *Contact and conflict in intergroup encounters* (pp. 1–44). Oxford: Blackwell.

Horenczyk, G. (1996). Migrant identities in conflict: Acculturation attitudes and perceived acculturation ideologies. In M. Breakwell & E. Lyons (Eds.), *Changing European Identities: Social psychological analyses of social change* (pp. 241–250). Oxford: Butterworth-Heinemann.

Hornsey, M. J., & Hogg, M. A. (2000). Assimilation and diversity: An integrative model of subgroup relations. *Personality and Social Psychology Review*, *4*, 143–156.

Huo, Y. J., & Molina, L. E. (2006). Is pluralism a viable model of diversity? The benefits and limits of subgroup respect. *Group Processes and Intergroup Relations*, *9*, 359–376.

Huo, Y. J., & Tyler, T. R. (2001). Ethnic diversity and the viability of organizations: The role of procedural justice in bridging differences. In J. Greenberg & R. Cropanzano (Eds.), *Advances ion organization justice* (pp. 213–244). Stanford, CA: Stanford University Press.

Jetten, J., Spears, R., & Manstead, A. S. R. (1997). Strength of identification and intergroup differentiation: The influence of group norms. *European Journal of Social Psychology*, *27*, 8161–8167.

Klonoff, E. A., Landrine, H., & Campbell, R. (2000). Sexist discrimination may account for well-known gender differences in psychiatric symptoms. *Psychology of Women Quarterly*, *24*, 93–99.

Koper, G., Van Knippenberg, D., Bouhuijs, F., Vermunt, R., & Wilke, H. (1993). Procedural fairness and self-esteem. *European Journal of Social Psychology*, *23*, 313–325.

Kray, L. J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Experimental Social Psychology*, *12*, 942–958.

Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York: McGraw-Hill.

Major, B., Quinton, W. J., & Schmader, T. (2003). Attributions to discrimination and self-esteem: Impact of group identification and situational ambiguity. *Journal of Experimental Social Psychology*, *39*, 220–231.

McAllister, D. J., & Bigley, G. A. (2002). Work context and the (re)definition of self: How organizational care influences organization-based self-esteem. *Academy of Management Journal*, *45*, 894–904.

Moya, M., Glick, P., Exposito, P., & Casado, P. (2007). *It's for your own good: Women's tolerance of benevolently-justified discrimination*. Granada: University of Granada, Unpublished manuscript.

Oakes, P. J., Haslam, S. A., & Turner, J. C. (1994). *Stereotyping and social reality*. Oxford: Blackwell.

Ouwerkerk, J. W., de Gilder, D., & de Vries, N. K. (2000). When the going gets tough, the tough get going: Social identification and individual effort in intergroup competition. *Personality and Social Psychology Bulletin*, *26*, 1550–1559.

Pierce, J. L., Gardner, D. G., Cummins, L. L., & Dunham, R. B. (1989). Organizational-based self-esteem: Construct definition measurement and validation. *Academy of Management Journal*, *32*, 622–648.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*, 336–356.

Schopler, J., & Insko, C. A. (1992). The discontinuity effect: Generality and mediation. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (pp. 121–151). Chichester: Wiley.

Spears, R., Doosje, B., & Ellemers, N. (1997). Self-stereotyping in the face of threats to group status and distinctiveness: The role of group identification. *Personality and Social Psychology Bulletin*, *23*, 538–553.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379–440). San Diego, CA: Academic Press.

Swann, W. B. (1990). To be adored or to be known: The interplay of self-enhancement and self-verification. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (Vol. 2, pp. 408–448). New York: Guilford Press.

Swann, W. B., Jr., & Ely, R. J. (1984). A battle of wills: Self-verification versus behavioral confirmation. *Journal of Personality and Social Psychology*, *46*, 1287–1302.

Swann, W. B., & Read, S. J. (1981). Self-verification: How we sustain our self-conception. *Journal of Experimental Social Psychology*, *17*, 351–372.

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*, 199–214.

Tajfel, H. (1981). The social psychology of minorities. In H. Tajfel (Ed.), *Human groups and social categories: Studies in social psychology*. Cambridge: Cambridge University Press.

Terry, D. J., Carey, C. J., & Callan, V. J. (2001). Employee adjustment to an organizational merger: An intergroup perspective. *Personality and Social Psychology Bulletin*, *27*, 267–280.

Truax, K., Cordova, D. I., Wood, A., Wright, E., & Crosby, F. (1998). Undermined? Affirmative action from the targets' point of view. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The target's perspective*. San Diego, CA: Academic Press.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorisation theory*. Oxford: Blackwell.

Tyler, T. R., & Blader, S. L. (2002). Terms of engagement: Why do people invest themselves in work? In H. Sondak (Ed.), *Towards a phenomenology of groups and group membership* (pp. 115–140). New York: Elsevier Science.

Tyler, T., Degoey, P., & Smith, H. (1996). Understanding why the justice of group procedure matters: A test of the psychological dynamics of the group-value model. *Journal of Personality and Social Psychology*, *70*, 913–930.

Van den Bos, K., Wilke, H. A. M., Lind, E. A., & Vermunt, R. (1998). Evaluating outcomes by means of the fair process effect: Evidence for different processes in fairness and satisfaction judgments. *Journal of Personality and Social Psychology*, *74*, 1493–1503.

Van Oudenhoven, J. P., Prins, K., & Buunk, B. (1998). Attitudes of minority and majority members towards adaptation of immigrants. *European Journal of Social Psychology*, *28*, 995–1013.

Received 21 January 2008; revised version received 2 December 2008