

## Who is this Donald? How social categorization affects aggression-priming effects

SABINE OTTEN<sup>1\*</sup> AND DIEDERIK A. STAPEL<sup>2</sup>

<sup>1</sup>University of Groningen, The Netherlands

<sup>2</sup>Tilburg University, The Netherlands

### Abstract

*When investigating the effects of aggression on person perception, many studies have relied on the famous 'Donald' paradigm (e.g., Srull & Wyer, 1979). Little attention has been paid, however, to the role that is played by the social identity of Donald (the target of perception) in such effects. Three studies with Dutch participants, consistently reveal that judgments of an ingroup (Dutch) target are less prone to be affected by aggression priming than judgments of an outgroup (Moroccan, Surinamese, German) target, and that this effect is moderated by the extent to which the target's group is associated with aggressiveness. Importantly, the studies show that such association can be created not only via well-established stereotypes (e.g., 'Moroccans are aggressive'), but also via subtle contextual priming. When priming activates an association between ingroup and aggressiveness ('The Dutch are aggressive'), the subsequent evaluation of the ingroup targets can also be influenced negatively. The present research thus demonstrates that social categorizations and the contextual associations attached to them delineate an important boundary condition for aggression priming effects. Copyright © 2006 John Wiley & Sons, Ltd.*

An important condition for behavior to be defined as aggressive is the perception of an intention to do harm (see Baron, 1979). Often, however, hostile intention cannot be derived unequivocally from behavior and interpreting behavior as aggressive is thus a matter of interpretation (see Dodge & Somberg, 1987; Mummendey, Linneweber, & Löschper, 1984). Furthermore, as decades of social psychological 'priming' research have shown, the extent to which someone will interpret behavior as *aggressive* instead of in terms of other applicable concepts (such as *playful, assertive, angry, frustrated*) is determined by the cognitive *accessibility* of these concepts (Srull & Wyer, 1979). Primed information is more likely to exert an effect on the interpretation of behavior than non-primed information, *even* when people are not consciously aware of this information (see e.g., Devine, 1989).

\*Correspondence to: Dr Sabine Otten, University of Groningen, Department of Social and Organizational Psychology, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands. E-mail: s.otten@rug.nl

Interestingly, so far there has been little research about the role of social categorization variables in the occurrence and magnitude of priming effects on person judgments (for exceptions, see Banaji, Hardin, & Rothman, 1993; Wigboldus, Dijksterhuis, & van Knippenberg, 2003). This is surprising, especially in the context of the interpretation of hostile or aggressive behavior. After all, 'seeing' hostility in a specific act may heavily depend on the social category of the actor, such that the behavior of 'outgroup' members is more readily interpreted in terms of aggressiveness than the behavior of 'ingroup' members. In the present paper, we will fill this void in the research literature and analyze the interaction between accessibility and intergroup processes. That is, we will investigate how 'priming' aggressiveness may guide the interpretation of ambiguously aggressive behavior of ingroup and outgroup targets. First, we assume that there is a general tendency to favor ingroup over outgroup targets (Hewstone, Rubin, & Willis, 2002), such that ambiguously behaving ingroup targets are less willingly described as aggressive than outgroup targets showing the same behavior. Second, and more importantly, we hypothesize that such intergroup bias in aggression priming effects may be moderated not only by chronically accessible stereotypes, but also by situationally activated associations between social categories and aggressiveness.

### PRIMING THE TARGET

When the concept of aggression is cognitively accessible, people will more readily see hostile intent in ambiguously aggressive behavior. Thus, when you walk home from the movie theatre you are more likely to expect an aggressive attack when another person is quickly approaching you in a dark street when the movie you have just seen is *The Texas Chainsaw Massacre* rather than *Terms of Endearment*. Similar accessibility-driven interpretation effects may be obtained when people are exposed to social category information that is stereotypically associated with the trait 'aggressiveness.' Thus, as several classic stereotyping studies have shown, people are more likely to interpret ambiguously hostile behavior in terms of aggressiveness when they have previously been primed with groups such as 'Blacks' or 'Hooligans' or 'Skinheads,' because these groups are associated with the trait 'aggressiveness' and thus likely to spontaneously activate this particular trait (see e.g., Devine, 1989; Dovidio, Evans, & Tyler, 1986).

Previous priming research has thus provided strong evidence that when contexts and stereotypes activate aggressiveness-related information they are likely to guide the interpretation of ambiguously aggressive behavior. However, to our knowledge, most of the emphasis in previous aggression-priming research has been on what is *activated* rather than on what is to be *perceived* or *judged*. That is, the role of the 'target' has been largely neglected. Hence, what is missing is a systematic, empirical analysis of the role of the target in aggression-priming effects (for a similar analysis see Banaji et al., 1993; Higgins, 1996; Wigboldus et al., 2003).

The present research hopes to fill this gap by investigating the impact of the social categorization of the target on the perception of hostile intent. More specifically, we want to investigate whether the *mere* categorization of the target as either an ingroup or an outgroup member affects the interpretation of ambiguously aggressive behavior. Is the behavior of an outgroup member more readily interpreted in terms of aggression than that of an ingroup member? Even when there is no stereotype to warrant such an interpretation? Or are such aggression-interpretations only likely to occur when the target person is a member of a stereotypically aggressive outgroup? Furthermore, we are interested how such social categorization effects *interact* with aggression-concept priming. Will the effects of aggression-concept priming dwindle or disappear when the person to be judged is an ingroup member? Those are the

questions we want to address by investigating the impact of concept and stereotype priming in an intergroup context.

### WHO IS THIS DONALD?

In one of the first aggression priming studies, Srull and Wyer (1979) demonstrated that perceptions of hostile intent can be activated relatively subtly and implicitly by using what became to be known as the 'Donald paradigm.' In this paradigm, participants first had to unscramble a series of short sentences, some of which dealt with aggressive behavior (e.g., 'break his leg' see Costin, 1969). Then, in an ostensibly unrelated second study, participants were given an ambiguously aggressive description of 'Donald' and asked to judge Donald on a number of traits. Srull and Wyer (1979) found that, compared to a control condition (in which neutral sentences had to be unscrambled), participants viewed Donald as relatively aggressive and hostile. Srull and Wyer's 'Donald' has been used many times, to demonstrate a large number of aggression-priming effects. For example, Bargh and Pietromonaco (1982) used Donald in their studies of subliminal priming effects, Devine (1989) used Donald in her study of automatic stereotyping effects, and Herr (1986) used Donald in his study of assimilation and contrast effects (see also Stapel, Koomen, & van der Pligt, 1997).

Interestingly, in most of these Donald-studies (as well as in studies that used slightly different priming paradigms), little attention has been paid to a factor that is known to be highly relevant for the interpretation of behavior, namely the *social categorization* of the target person. Who is this 'Donald' after all? Is he one of *us* or one of *them*? Is he an ingroup or an outgroup member? Is the group he belongs to stereotyped as aggressive? Ample research on intergroup behavior suggests that the answers to these questions provide important determinants of the interpretation of people's behavior. For example, ingroup members are typically seen in more positive terms than outgroup members (see Hewstone et al., 2002). Such a positivity bias should *decrease* perceived hostility when evaluating an ambiguously aggressive ingroup member and *increase* perceived hostility when evaluating an ambiguously aggressive outgroup member.

However, a study by Banaji et al. (1993) may be interpreted as suggesting that in the Donald paradigm such intergroup processes and self-categorization biases are relatively less likely to have an effect. These authors studied the impact of aggression-concept priming on judgments of an ambiguously aggressive target person, named 'Donald' (a man) or 'Donna' (a woman). Results showed that aggression-concept priming only affected judgments of Donald, but not Donna, presumably because aggressiveness is part and parcel of the stereotype of males but not of females (men are aggressive, women are not). Interestingly, in the Banaji et al. (1993) study participants' gender had no effect on judgments. Thus, strictly speaking, whether the target was an ingroup member (Donald for men, Donna for women) or an outgroup member (Donna for men, Donald for women) did not affect participants' judgments. Does this mean that social categorization does not play a role in the Donald paradigm, and that group-serving biases protecting the collective self ('Donald is one of us: he is not really aggressive') are unlikely to occur in that paradigm?

We do not think so. We believe that it would be inappropriate to conclude from the Banaji et al. (1993) study that ingroup-outgroup distinctions do not matter in this domain of research. In fact, as Banaji et al. themselves indicate, using gender as a social category in this domain of research is a very specific case, as most men and women know these stereotypes (men are aggressive, women are not), consider them a fact of life, and are hardly offended by them (Banaji et al., 1993; Stapel & Koomen, 1998; Swim, 1994). In a context, however, where there is more intergroup segregation and conflict, and where the link between categorization and aggression is subject to dissent (e.g., African-Americans

may not agree that they are more aggressive than European-Americans), there are more likely to be positivity biases in the perception of the ambiguously hostile behavior of ingroup members and negativity biases in the perception of the ambiguously hostile behavior of outgroup members. In such a case, the mere categorization of a target person as either an ingroup member or an outgroup member may determine the strength of aggression-priming effects.

Finally, we want to stress that our research, though similar with respect to its interest in social categorization as powerful contextual variable affecting automatic judgmental processes, goes beyond findings reported by Wigboldus and collaborators (2003). In line with Kunda and Thargard (1996), they showed in a series of studies that social categorization can affect impression formation by inhibiting spontaneous trait-inferences that do not fit the stereotype associated with the target category. For example, the sentence 'X put the found money in his own purse' spontaneously implies the trait 'dishonest' if the actor in the sentence is specified as 'a junkie,' whereas the same inference is inhibited when the actor happens to be a priest. In our own research, we also assume that stereotype-consistency determines whether aggression priming will affect subsequent person perception. Yet, we do not only test this idea in a different paradigm. More importantly, we investigate whether (a) the ingroup-outgroup distinction *per se* affects impression formation, and whether (b) not only chronically accessible stereotypes, but also ad hoc, situationally activated, associations between the target's group and aggressiveness can moderate aggression priming effects, thereby demonstrating the malleability of social categorization effects.

## SUMMARY AND OVERVIEW

The present research will focus on the interplay between priming (the activation of the concept of aggression) and social categorization effects. Specifically, we will investigate how the cognitive activation of the concept of aggression affects the interpretation of ambiguously aggressive behavior with special attention for the social category of the target person engaging in this behavior (Is the target an ingroup or an outgroup member? Is the target's group stereotypically viewed as aggressive?). Thus, the present studies should further our understanding of the role of cognitive accessibility on the perception of hostile intent as well as further delineate the boundary conditions of priming effects: we set to demonstrate that by subtly indicating who 'Donald' is (ingroup member or outgroup member; member of a stereotypically aggressive or not stereotypically aggressive group), the impact of primed concepts may change dramatically.

In the present studies, we use Srull and Wyer's (1979) Donald paradigm to investigate the interaction between priming and social categorization effects in the interpretation of ambiguous aggressive behavior. Extending previous Donald studies, we will vary the group affiliation of the target person, such that 'Donald' is either an ingroup member, a member of a stereotypically aggressive outgroup, or a member of an outgroup that is not seen as markedly aggressive.

We predict that in control priming conditions, the target's social categorization will determine how the target is judged, such that group stereotypes, and, possibly, ingroup-favoring tendencies, will guide the interpretation of ambiguous behavior. When the concept of aggression is primed, on average, the target should be perceived as relatively more hostile. This effect, however, should be stronger for a target person that is outgroup rather than ingroup member.

In addition, we assume that this bias is moderated by category-specific chronic or situational differences in stereotype applicability. If the target is a member of a stereotypically aggressive outgroup (such as Moroccans), ambiguous behavior is more likely to be viewed as aggressive than if the target belongs to a stereotypically non-aggressive group (the Surinamese, the Dutch). However, when our participants (Dutch students) have just learned (i.e., have just been 'primed' with information) that their

own group (Dutch students) has become quite violent recently, then an ingroup target may be viewed as relatively even more aggressive than an outgroup target. Hence, we expect—and test in studies 2 and 3—an interaction between type of priming and target association such that the category involved in the aggression priming (e.g., ‘the Dutch are aggressive’ vs. ‘the Moroccans are aggressive’) will determine which target (ingroup or outgroup) will be perceived as more aggressive.

In a pilot study we validated the assumed differences in the stereotypicality of aggressiveness as trait describing the relevant groups in the present studies (the Dutch, the Surinamese, the Moroccans, and the Germans). In Study 1, we compared priming effects for ‘Siebe’, a Dutch target (ingroup), ‘Hafid’, a Moroccan target (outgroup, stereotypically seen as aggressive), and ‘Patrick’, a Surinamese target (outgroup, not stereotyped as aggressive). We hypothesize that the effect of aggression priming as compared to a control condition without priming should be moderated by target categorization: Priming effects should be stronger for the two outgroup conditions than for the ingroup condition. Moreover, in line with the pilot results, we expected that a Moroccan target *per se* will already activate (‘prime’) aggressiveness in the control condition, whereas the same should not apply to a Surinamese target. Aggression priming, however, should affect judgments of both outgroup targets.

In Study 2, we vary stereotype applicability not only for the outgroup, but also for the ingroup. Rather than relying on chronically accessible stereotypes, we manipulate *situationally* how far ingroup and outgroup were associated with aggressiveness. We hypothesize an interaction effect such that when combining aggression priming with a specific category (Dutch/ingroup or Moroccan/outgroup), aggression-priming effect should only occur if primed category and target category match. In Study 3, finally, we replicate Study 2 with a different outgroup (Germans). Herewith, we do not only intend to demonstrate the reliability of our findings, but also to show that our results are not restricted to ingroup–outgroup comparisons characterized by asymmetrical status relations.

## PILOT STUDY

The four relevant social categories in the main studies reported in the following are the Dutch, the Surinamese, the Moroccans, and the Germans. In our research we did not only want to show effects of target categorization stemming from the activation of stereotypes associated with the respective social categories. We also wanted to provide evidence for a more general ingroup–outgroup bias in aggression priming effects on subsequent person perception. To this end, we needed to compare the evaluations of targets from either a stereotypically aggressive outgroup, and from an outgroup that, in general, is not associated considerably more strongly with aggressiveness than the Dutch ingroup. Previous research on stereotypes about ethnic groups in the Netherlands revealed that Moroccans are typically seen aggressive, whereas this does not apply to the Surinamese (Gordijn, Koomen, & Stapel, 2001); moreover, Oldenhuis, Gordijn, and Otten (2005) found that with respect to aggressiveness, the Dutch place Germans somewhat in between Moroccans and the Surinamese. In both of these studies, stereotypes were assessed by open-ended questions (‘Name a couple of traits you consider typical for ...’). In our pilot study, we wanted to validate these findings in a closed response format and with the Dutch as comparison standard.

### Method

Participants were 50 Dutch students. These students filled in a questionnaire in which the Dutch were compared to Moroccans, to Surinamese, and to Germans on a total of 18 trait dimensions,<sup>1</sup> including

<sup>1</sup>We used two versions of the questionnaire, varying the sequence of the three ingroup–outgroup comparisons.

the here relevant traits of aggressiveness and hostility. Participants indicated for each trait on a 7-point scale whether, to their opinion, it applied more to the Dutch or to the Moroccans (and Surinamese or German, respectively). A score of '1' implied that the trait solely applied to the Dutch, a score of '7' that it solely applied to the respective comparison outgroup. Besides the trait ratings, we also measured overall prejudice ('How positive or negative do the Dutch, on average, see the Moroccans (Surinamese/Germans)?', 1 = positive, 7 = negative) and on the perceived relative status of the outgroup ('Compared to the indigenous Dutch, how do you perceive the status of Moroccans in the Netherlands?' 1 = much lower, 7 = much higher).

## Results and Discussion

A repeated measures MANOVA with type of comparison (Dutch–Surinamese, Dutch–German, Dutch–Moroccan) as between-participants variable revealed for both relevant traits a significant effect; aggressiveness:  $F = 27.05$ ,  $p < 0.001$ ,  $\eta^2 = 0.36$ , hostility:  $F = 17.52$ ,  $p < 0.001$ ,  $\eta^2 = 0.27$ . With respect to aggressiveness, the score for the Dutch–Surinamese comparison was lowest ( $M = 4.30$ ); indicating that aggressiveness was nearly as little assigned to the Surinamese as it was assigned to the Dutch), followed by the relative score for the Germans ( $M = 4.66$ ), and then the Moroccans ( $M = 5.76$ ). Simple comparisons revealed that the scores for Germans and Surinamese did not differ significantly, whereas the score for the Moroccans differed from the two others with  $t(49) = 6.38$  and  $6.68$ , respectively,  $ps < 0.001$ . The results for the ratings on hostility reveal a similar picture; the scores were lowest for the Surinamese ( $M = 4.22$ ), followed by the Germans ( $M = 4.82$ ) and the Moroccans ( $M = 5.44$ ). All simple comparisons were significant,  $3.16 < t < 5.75$ , all  $ps < 0.003$ . Moreover, with respect to global prejudice, the Surinamese ( $M = 3.94$ ) scored significantly better than the Germans ( $M = 4.50$ ), who scored significantly better than the Moroccans ( $M = 5.76$ ). All paired comparisons were significant,  $2.33 < t < 8.05$ , all  $ps < 0.05$ . Finally, the Moroccans ( $M = 2.65$ ) were seen as having a markedly lower status than the (indigenous) Dutch, whereas the same did not apply to the German and the Surinamese outgroup ( $M = 4.68$ , and  $M = 4.50$ ). Whereas the scores for Germans and the Surinamese do not differ,  $t < 1$ , the score for the Moroccan outgroup is significantly lower,  $t = 9.16$  and  $t = 8.65$ , respectively,  $ps < 0.001$ ).

## STUDY 1

In this first study, we investigated the effects of target categorization on aggression priming effects. We predict these effects to be stronger for outgroup than for ingroup targets. Moreover, we hypothesize an additive effect of aggression priming and of chronically accessible stereotypes on the perceptions of outgroup targets as aggressive.

## Method

### *Participants*

One hundred forty Dutch students (all first year-students in Psychology) participated. They were randomly assigned to the conditions of a 2 (Prime: aggression, neutral)  $\times$  3 (Target: Dutch/ingroup, Surinamese/non-aggressive outgroup, Moroccan/aggressive outgroup) between-participants design.

### *Priming Task*

The experiment was part of a general testing session in which participants received several questionnaires. The priming task was entitled 'Language Comprehension.' This task consisted of a page of 26 scrambled four-or five-word groups (He dog hits the telephone) that participants had to unscramble by deleting one word. In the experimental conditions, 14 of these sentences described aggressive behaviors and 12 described neutral behaviors. In the neutral condition all 26 sentences were neutral. Participants' task was to reorganize the word groups into meaningful sentences. The priming stimuli were Dutch translations of the material that was developed by Srull and Wyer (1979), see also Stapel et al., 1997).

### *Target Description and Measures*

After participants had finished the priming task and a filler task (identifying 16 neutral words, e.g., street, room, new, out of a word puzzle, consisting of a matrix of  $5 \times 28$  letters), they were instructed to put the 'language comprehension' booklet in a folder on their desks. Next, they were given the impression formation booklets, entitled 'Impression Formation.' They were instructed to read the target paragraph and to try to form an impression of the target. The target paragraph described a series of activities involving the target to be judged and included behaviors only ambiguously related to aggressiveness among neutral behavior (see Srull & Wyer, 1979; Stapel & Koomen, 1998). In the *Dutch target* conditions the target was named 'Siebe, born in The Netherlands.' In the *Surinamese target* conditions, the target was named 'Patrick, born in Surinam.' In the *Moroccan target* conditions, the target was named 'Hafid, born in Morocco.'

Participants were asked to rate the target (Siebe, Patrick, Hafid) on eight unipolar trait dimensions adapted from Srull and Wyer (1979). Participants indicated their impressions by scoring three trait dimensions that implied a high degree of hostility ('hostile,' 'aggressive,' 'unfriendly') and five trait dimensions that were unrelated to hostility ('selfish,' 'fretful,' 'intelligent,' 'dependable,' 'helpful'). We included the latter scales to decrease the possibility that participants would become suspicious that the concept of interest was hostility-related. Related and unrelated rating scales were interspersed with each other. Ratings were made along a scale from 1 ('not at all') to 9 ('extremely').

Neither in Study 1 nor in studies 2 and 3 did we find main or interaction effects on the unrelated rating scales ( $F_s < 1$ ). Therefore, we further ignore these scales in the results sections of the current studies.

On the final page of the impression formation booklet, participants were probed as to what they thought the study was about. Specifically, participants were asked (a) in an open-ended response format what they thought the purpose of the study had been, and two questions in a dichotomous response format (yes vs. no), namely (b) whether they thought the language comprehension and the impression formations tasks were related and (c) whether they thought that their answers in impression formation task were affected by the language comprehension task (for similar procedures see e.g., Stapel & Koomen, 1998). None of the participants showed any suspicion or awareness of a relation between the two tasks. On completion of the booklet, participants were debriefed and thanked for their time.

## **Results and Discussion**

Table 1 shows participants' mean ratings on the hostility scale (Cronbach's  $\alpha = 0.84$ ) as a function of the manipulations. An analysis of variance (ANOVA) revealed a Prime  $\times$  Target interaction,

Table 1. Priming ('aggressive' vs. control)  $\times$  Target (Surinamese, Moroccan, or Dutch); means and standard deviations for target's aggressiveness ratings (Study 1)

Target's aggressiveness	Target:		
	Surinamese Patrick	Moroccan Hafid	Dutch Siebe
Prime			
Aggressive	5.55 <sup>b</sup> (1.03)	6.39 <sup>c</sup> (1.06)	4.33 <sup>a</sup> (1.16)
Control	4.33 <sup>a</sup> (1.47)	5.46 <sup>b</sup> (1.75)	4.60 <sup>a</sup> (1.87)

Note: Lower score mean less negative judgments, on a compound scale comprising the mean scores from three aggression-related items (9 = maximum). All means with different superscripts differ according to *post hoc* simple comparisons with  $p < 0.05$ , except the comparison between Dutch Siebe and Moroccan Hafid in the control condition, which differs with  $p = 0.06$ .

$F(2, 134) = 3.19, p < 0.05, \eta^2 = 0.05$ , a main effect for Target,  $F(2, 134) = 11.44, p < 0.001, \eta^2 = 0.15$ , and a main effect of Prime,  $F(1, 134) = 6.01, p < 0.05, \eta^2 = 0.04$ . As can be seen in Table 1, these effects reveal that aggression priming affected judgments of an outgroup target but had no effect on ingroup target judgments, ( $F < 1$ ). Furthermore, the aggressively stereotyped outgroup target (Moroccan) is seen as relatively more aggressive than the other two targets, even when not primed. Judgments of the non-aggressive outgroup target (Surinamese), however, do not show this effect. In the control priming condition, judgments of the Dutch and the Surinamese target do not differ ( $F < 1$ ).

The results of Study 1 clearly confirm our hypotheses. First, aggression priming increases the likelihood that ambiguously hostile behavior is interpreted in accordance with this prime. However, the present findings indicate that there is an important boundary to this well-known priming effect: Aggression priming only affects judgments of an outgroup target. Judgments of ingroup targets are unaffected by aggression primes. Interestingly, this ingroup-outgroup effect revealed itself independent of whether the outgroup target belonged to a group that is stereotypically associated with aggressiveness (Moroccans) or to an outgroup to which this stereotype does not apply (Surinamese). This suggests that the distinction between 'us' versus 'them' is sufficient to guide aggression-priming effects.

These findings show that whereas outgroup ratings were clearly (and negatively) affected by the aggression priming, participants did not use activated and applicable trait information in their judgments of an ingroup member, thus evaluating ingroup members more positively (see Brewer & Brown, 1998; Hewstone, Rubin, & Willis, 2002; Otten, 2002). Moreover, chronically accessible stereotypes ('Moroccans are aggressive') additionally affected the strength of priming effects: Overall, the Moroccan target received the highest aggression ratings.

## STUDY 2

In Study 1, we demonstrated that ingroup-outgroup processes and existing stereotypes can both amplify and weaken well-known priming effects. In the present study we investigate to what extent situationally created rather than chronically existing stereotypes may moderate priming effects. Other than Moroccans, the Dutch, the ingroup in our research, are typically not stereotyped as 'aggressive.' However, this can be contextually changed by subtly priming an association between this ingroup and aggressive behavior. When subtly exposing (Dutch) participants to information that their ingroup is aggressive, it should be possible to change the pattern of results found in our first study.

Study 2 investigates the effects of target categorization (ingroup, outgroup) in combination with an experimental manipulation of the fit between the primed trait and the target's group membership ('The Dutch are aggressive'; 'Moroccans are aggressive'). We expect to replicate the findings of Study 1 when linking the concept aggression with the outgroup label: In this case, the ingroup target ('Siebe') should be evaluated as less hostile and aggressive than the outgroup target ('Hafid'). This effect, however, should vanish or even reverse when priming aggression together with the ingroup label.

## Method

### *Participants*

Hundred Dutch first-year psychology students participated in return for course-credit. They were randomly assigned to one of the four conditions of the study. Again, the two dichotomous questions and the one open-ended question probing participants' suspicion or awareness (see Study 1) did reveal that all participants were blind with respect to the manipulation and the true purpose of the experiment.

### *Design*

The design was a 2 (Priming: aggression + Dutch/ingroup, aggression + Moroccan/outgroup)  $\times$  2 (Target: Dutch/ingroup, Moroccan/outgroup) between-participants design.<sup>2</sup> The dependent measures critical to our hypotheses were the ratings of the target on the trait dimensions referring to hostility and aggressiveness (Cronbach's alpha = 0.73).

### *Procedure*

The procedure was very similar to the one used in Study 1. The only difference was that the four critical sentences that were supposed to prime aggression were now adapted such that they additionally primed a group label (ingroup or outgroup), for example, 'The Dutchman loses his temper.'

## Results and Discussion

We ran a 2  $\times$  2-between-subjects ANOVA with mean aggressiveness as dependent variable. The analysis revealed the expected Priming  $\times$  Target interaction,  $F(1, 96) = 43.31, p < 0.001, \eta^2 = 0.31$ , and a main effect for Target,  $F(1, 96) = 11.01, p = 0.001, \eta^2 = 0.11$ . The main effect of Prime was not significant,  $F < 1$ . As can be seen in Table 2, there is a crossover-interaction such that when the group label included in the aggression priming fits the target categorization, hostility ratings are significantly higher than when there was no match between the primed group label and target. Importantly, this pattern similarly holds for both ratings about ingroup targets ('Siebe') and outgroup targets ('Hafid').

Hence, on the one hand, if the aggression priming relates to the Dutch, then a Dutch (ingroup) target is rated as relatively aggressive and a Moroccan target is rated as relatively non-aggressive, even though

<sup>2</sup>In Study 2, we manipulated target only as a two-level factor, contrasting reactions to either a Dutch or a Moroccan target. Hereby, we compared the ingroup-target with a target whose category already, per default, is associated with aggressiveness. This implies an especially conservative test for the hypothesized context-specific reversal of the effects of target categorization on aggression priming.

Table 2. 'Group label'-aggression priming ('Moroccans are aggressive' vs. 'the Dutch are aggressive') × Target (Moroccan vs. Dutch); means and standard deviations for target's aggressiveness ratings (Study 2)

Target's aggressiveness	Target	
	Moroccan Hafid	Dutch Siebe
Prime		
Moroccans are aggressive	5.67 <sup>d</sup> (1.03)	3.43 <sup>a</sup> (1.10)
Dutch are aggressive	4.36 <sup>b</sup> (1.16)	5.06 <sup>c</sup> (1.17)

Note: Lower score mean less negative judgments, on a compound scale comprising the mean scores from three aggression-related items (9 = maximum). All means with different superscripts differ according to *post hoc* simple comparisons with  $p < 0.05$ , except the target main effect within the 'Dutch are aggressive'-priming condition, which differs with  $p = 0.07$ .

the group of Moroccans is chronically stereotyped as aggressive. If, on the other hand, the aggression priming includes the Moroccan group label, then a Moroccan target is seen as aggressive and a Dutch target receives low ratings on this trait dimension.

This pattern of results impressively documents that even though there seems to be a buffer protecting ingroup targets against aggression priming effects (Study 1), this buffer can easily be overcome by situationally priming links between the ingroup and aggressiveness. Obviously, seeing an ingroup member as *not* aggressive, as would be functional for the striving for positive social identities (Tajfel & Turner, 1986), is not unconditional. If the applicability of the aggression concept is enhanced by linking it to the ingroup, then the ambiguous description of an ingroup target can also be interpreted in terms of aggressiveness.

### STUDY 3

Our third study served two main purposes. First and foremost, we wanted to replicate the findings of Study 2 demonstrating that ingroup bias in aggression priming effects can be reduced and even reversed by situationally priming not only the concept of aggression, but also an association of the ingroup with aggressiveness. To investigate the generalizability of our findings, in this experiment we used a different outgroup. The target in the outgroup condition was German rather than Moroccan. As shown in the pilot study, this outgroup is much less strongly than Moroccans stereotypically associated with aggressiveness. At the same time, the Germans are seen as higher in status than Moroccans. We expected to replicate the findings from Study 2, such that depending on which group label was linked to the aggression priming the respective target evaluation would be assimilated to the priming. Second, in Study 3 we included not only a control condition without priming (as in Study 1), but also a hanging control where the target was not explicitly categorized as ingroup or outgroup. Herewith, we can address the question as to why in Study 1 we did not find any evidence for aggression priming effects on the ingroup target. This seems to contradict the ample evidence from the Donald paradigm revealing priming effects for 'Donald,' a name more popular amongst Whites, who mostly were the participants in the classical experiments with this paradigm (e.g., Bargh & Pietromonaco, 1982; Srull & Wyer, 1979). We assume that the explicit categorization of the ingroup target ('Siebe from the Netherlands') is responsible here. Through the explicit categorization as ingroup member, the target becomes 'one of us,' and as a default (e.g., Otten & Moskowitz, 2000), its evaluation becomes more positive. Therefore, we assume that a target categorized as 'from the Netherlands' will be evaluated more positively than an

uncategorized target. We called our uncategorized target 'Erik,' which is a quite common name in the Netherlands, but which—other than 'Siebe'—also exists in other countries (including Germany).

## Method

### *Participants and Design*

One hundred and eighty-six students from a Dutch university participated in the experiment. They received course credit in exchange for their participation. Participants were assigned at random to one of the experimental conditions of a  $3 \times 2$  design with Priming (aggression plus ingroup, aggression plus outgroup, control) and Target (Siebe/ingroup, Dieter/outgroup) and an additional hanging control (no aggression priming, Erik/uncategorized target).

### *Method*

The method in this study was exactly identical to Study 2. The only deviation was that now the Germans were the relevant outgroup and 'Dieter' the outgroup target, respectively. Moreover, we included a control condition with 'Erik' (a name that *can* be Dutch, but also German or Belgian, etc.) as a target, and without adding any further categorical information to this name. Besides, we extended the check as to whether the participants were blind to the experimental conditions: In addition to the one open-ended and two dichotomous questions used in studies 1 and 2, we included another item asking whether or not participants thought that their evaluation of the target person was influenced by something other than reading the text about him. Again, we succeeded in hiding the true purpose of the study. Only five participants gave yes-responses to one or more of the three probing questions. Only two of them indicated that they expected a link between the sentence-unscrambling task and the person-perception task. However, their responses to the question about which link they assumed there was, revealed that these participants were also blind to the real experimental conditions ('There is always a link between everything'; 'it just feels as if there is a link'). Therefore, we decided to keep the full sample for the subsequent analyses.<sup>3</sup>

## Results and Discussion

We expected to replicate Study 2 such that there would be an interaction effect between Priming (The Dutch are aggressive; the Germans are aggressive; control) and Target condition (Dutch Siebe; German Dieter) on the compound aggressiveness scale (items aggressive, hostile, and unfriendly;  $\alpha = 0.69$ ). As expected, a  $2 \times 2$  between-participants MANOVA with mean aggressiveness as dependent and Priming (the Dutch are aggressive, the Germans are aggressive) and Target (Dutch 'Siebe,' German 'Dieter') as independent variables revealed a significant interaction effect,  $F(1, 99) = 68.19, p < 0.001, \eta^2 = 0.42$ , qualifying the significant Target main effect,  $F(1, 99) = 11.15, p = 0.001, \eta^2 = 0.10$ . Simple comparisons reveal that in the ingroup target condition, aggression priming combined with the ingroup label leads to markedly higher aggression ratings than aggression priming combined with the outgroup label,  $F(1, 99) = 21.35, p < 0.001$ . In the *outgroup target condition*, the picture is exactly reversed,  $F(1, 99) = 26.09, p < 0.001$ ; as can be seen from Table 3, the

<sup>3</sup>When doing the subsequently reported analyses without these participants, results are fully comparable.

Table 3. 'Group label'-aggression priming ('Germans are aggressive' vs. 'the Dutch are aggressive' vs. control)  $\times$  Target (German vs. Dutch (vs. uncategorized)); means and standard deviations for target's aggressiveness ratings (Study 3)

Target's aggressiveness	German Dieter	Target: Dutch Siebe	Uncategorized Erik
Prime			
Germans are aggressive	6.07 <sup>d</sup> (0.87)	3.60 <sup>a</sup> (1.06)	
The Dutch are aggressive	4.23 <sup>b</sup> (1.24)	5.28 <sup>c</sup> (1.42)	
Control	5.16 <sup>c</sup> (1.31)	4.42 <sup>b</sup> (1.27)	4.77 <sup>bc</sup> (1.07)

Note: Lower score mean less negative judgments, on a compound scale comprising the mean scores from three aggression-related items (9 = maximum). All means with different superscripts differ according to *post hoc* simple comparisons with  $p < 0.05$ .

target ratings for the control priming condition fall in between the two extremes per target condition. Hence, our data suggest that the fit between primed group label and target categorization determines whether aggression priming will result in assimilation or contrast in subsequent target perception.

In addition, we tested our assumption that the uncategorized target ('Erik') would be evaluated less positively than the obvious ingroup target ('Siebe'). A oneway-ANOVA (only on the control condition) with target (Siebe, Erik, Dieter) as independent variable supported this assumption. The target effect was marginally significant with  $F = 3.00$ ,  $p = 0.055$ ,  $\eta^2 = 0.07$ . As hypothesized, when no aggression was primed, Siebe was seen as less aggressive than Dieter,  $M = 4.42$  versus  $M = 5.16$ ,  $p < 0.02$ , whereas the score for Erik ( $M = 4.77$ ) fell in between these two means, but did not differ significantly from either of them.

## GENERAL DISCUSSION

In three studies we investigated how aggression priming guides the interpretation of ambiguously aggressive behavior of an ingroup or an outgroup member. Typically, research in this domain (e.g., Bargh & Pietromonaco, 1982; Stapel et al., 1997) does not explicitly categorize the target that is evaluated after people have been primed with 'aggressiveness' or 'hostility.' Thus, the robustness and generalizability of aggression priming studies to real life situations is unknown. In real life, the targets of social perception commonly belong to specific social groups: These targets of perception are members of 'us' or 'them' and these in- and outgroups are stereotypically associated with specific traits ('we are tolerant,' 'they are aggressive').

In the present studies we investigated how typical aggression priming effects may be moderated by the social categorization of the target of perception. Interestingly, whereas the mainstream priming literature suggests that priming effects are relatively strong and robust, the current studies suggest that the strength of aggression priming effects on judgments on ambiguously aggressive behavior importantly depend on *who* performs this behavior. Whether aggression priming affects actual judgments is determined by the social category the target person belongs to (ingroup, outgroup) as well as by the (chronically or situationally induced) stereotypes (aggressive, non-aggressive) that are associated with this social category.

Obviously then, aggression priming effects do not take place in a social vacuum. Rather, they are part and parcel of a social context that to a good deal, is shaped by social categorization. Though this statement seems self-evident, it has nonetheless received too little consideration, so far (see Spears,

Gordijn, Dijksterhuis, & Stapel, 2004, for a similar reasoning). Our findings reveal the necessity to take into account social categorization when investigating aggression-priming effects. Not only do we need to consider the categorization of the target whose evaluation might be affected by the priming, but the priming itself can also comprise associations with certain social categories. As consistently shown in both Study 2 and Study 3, when primed associations with aggression are firmly linked to a certain social category, then only evaluations of those targets that fit the priming will become subject of the typical assimilation effect in subsequent person perception.

Intergroup research has provided strong evidence for people's tendency to generate and to protect favorable images of their ingroup (for surveys see Brewer & Brown, 1998; Hewstone, Rubin, & Willis, 2002). Accordingly, there should also be more reluctance to see hostile intent in ambiguous behavior if it is performed by an ingroup member rather than by an outgroup member. However, our findings go beyond typical findings on the 'ultimate attribution error' (Pettigrew, 1979) and group serving attribution bias (Hewstone & Ward, 1985; Islam & Hewstone, 1993). In the current experiments, aggressiveness is activated subtly and implicitly (see Banaji et al., 1993). In neither of the three studies reported here, participants reported to be aware of the influence of the prime on their judgments. Yet, our results imply that the categorization of the target as ingroup member, and the positive affect automatically associated with ingroup categories (see Otten & Moskowitz, 2000; Otten & Wentura, 1999), can function as a buffer against an automatic aggression priming effect. Outgroup members, also those who are not stereotypically associated with aggressiveness (the Germans in Study 3, and especially the Surinamese in Study 1), are not protected by such buffer. Here, the typical enhancement of aggressiveness ratings after aggression priming can emerge.

At first glance, our finding that the categorization of the target as either ingroup or outgroup member moderates aggression priming effects may be viewed as contradicting previous findings by Banaji et al. (1993). In their research, after aggression priming male and female participants were similarly prone to see hostile intent in a male target ('Donald') rather than a female target ('Donna'); no additional effect of the target's status as either ingroup or outgroup member emerged. Hence, unlike our Dutch participants in Study 1, male participants in Banaji's study did not show any ingroup favoritism. We assume, however, that strong and consensual stereotypical associations between gender and aggressiveness (e.g., Swim, 1994) have overridden possible ingroup-outgroup differences in target perception (herewith paralleling the findings by Wigboldus et al., 2003). Therefore, in the present studies, we did not only compare how a member of a stereotypically aggressive outgroup (Hafid from Morocco) was evaluated as compared to the Dutch ingroup member (Siebe), but we also included targets from stereotypically rather non-aggressive outgroups (Patrick from Surinam and Dieter from Germany). Importantly, the control conditions in Study 1 (just as our pilot study) revealed exactly these differences in stereotypical associations with aggressiveness for the different target groups (Moroccan, Surinamese, or Dutch): Without aggression priming, the Moroccan target was seen as significantly more aggressive than both other targets, and aggressiveness ratings were similarly low for both the Dutch (ingroup) target and the Surinamese (outgroup) target. Only the latter, however, was affected by the aggression priming, while the ratings for the Dutch target remained unchanged. Accordingly, we feel confident that the differential priming effects in Study 1 do not just reflect a stereotyping effect but point to the more general role of ingroup-outgroup distinctions as a powerful contextual variable.

Banaji and collaborators (1993) interpreted their finding that a male but not a female target was affected by aggression priming in terms of 'stereotype' or 'social' applicability. The idea that social applicability is an important moderator for the occurrence of priming effects, and that there are varying degrees of fit between priming and target categorization, was pivotal for our studies 2 and 3. However, different from Banaji's research, and also different from the more recent work by Wigboldus et al. (2003), we did not just rely on already existing stereotypical assumptions about the aggressiveness of certain ethnic groups. Instead, in order to manipulate this fit for both ingroup and outgroup target, we

primed specific 'ingroup-is-aggressive' and 'outgroup-is-aggressive' knowledge by linking the typical aggression priming to a specific group label (ingroup or outgroup). The crossover interaction between target categorization and type of priming impressively demonstrates that the hidden activation of concepts such as aggression needs to be understood in a specific social context (see Spears et al., 2004): For *both* Moroccan and Dutch targets (and for German and Dutch targets in Study 3), only a match between target category and primed category led to aggression priming effects (i.e., pronounced ratings of the target's aggressiveness), whereas no priming effects emerged when primed category and target category mismatched.<sup>4</sup> Thus, once priming includes associations with a certain social category then the step from activation to application will only pertain to targets that fit the social category that was involved in the priming. This is especially noteworthy given the subtle and relatively implicit way in which we manipulated this 'fit' in the present research; nonetheless did the effects of the situational priming of aggression-stereotypes fully override the pattern that conventional stereotypes would have suggested.

Moreover, studies 2 and 3 show that, once the concept of aggressiveness is activated, people's greater reluctance to perceive hostile intent in an ingroup target's ambiguously aggressive behavior, vanishes when changing the context of the priming. Once the priming links aggressiveness with the ingroup label, an ingroup target is affected much more negatively by the priming than an outgroup member, even if the latter stems from a group that already stereotypically is associated with aggressiveness. Traditional research on intergroup biases already revealed that ingroup favoritism is sensitive to reality constraints, such that, for example, low status groups acknowledge the outgroup's superiority on status-related dimensions (e.g., Sachdev & Bourhis, 1987; 1991; see also Ellemers, van Rijkswijk, Roefs, & Simons, 1997). Our findings take this evidence a step further, showing that even on an implicit, unconscious level we can successfully block ingroup-favoring judgmental tendencies.

In conclusion then, the present research demonstrates that social categorization constitutes a powerful contextual boundary condition for aggression priming effects. The target's categorization as ingroup member can inhibit that implicitly activating aggressiveness carries over to perceptions of hostile intent. Moreover, when the priming itself is linked to certain social categories, it only affects targets from matching categories. Importantly, not only the (in)consistency of the primed associations with existing, well-established stereotypes (such as shown by Banaji et al., 1993, or by Wigboldus et al., 2003) determine the aggression priming effects, but also situationally primed associations between aggressiveness and certain social categories. Importantly, these effects are so strong that they fully override the ingroup favoring tendencies revealed by our first study. However, we think that these findings are not only a relevant contribution to the growing evidence that subtle contextual factors may influence automatic cognitive processes (e.g., Wittenbrink, Judd, & Park, 2001), but also to the literature on intergroup processes.

Relevant questions for future research are to investigate in how far the effects we reported generalize across different ingroup—outgroup categorizations, and whether members of minority and majority groups show the same or a different pattern of responses. Moreover, the question arises in how far trait valence is a relevant moderator. Research on ingroup-favoritism would suggest that priming a positive trait concept should much more easily affect subsequent person perception of ingroup rather than outgroup members. Finally, given that identifying hostile intent in another person's behavior is crucial for launching and fueling aggressive conflicts (Mummendey & Otten, 1989), these findings not only further our understanding of automatic priming effects, but should also prove useful in future aggression research.

<sup>4</sup>When comparing the means with those obtained in the control condition of Study 1, we might even recognize a tendency to contrast the target evaluation away from the primed concept: For both the Moroccan and the Dutch target, hostility ratings are especially low when primed category and target category mismatch.

## REFERENCES

- Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, *65*, 272–281.
- Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology*, *43*, 437–449.
- Baron, R. A. (1979). Effects of victim pain cues, victim's race, and level of prior instigations upon physical aggression. *Journal of Applied Social Psychology*, *9*, 110–114.
- Brewer, M. B., & Brown, R. J. (1998). Intergroup relations. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 554–594). New York: McGraw-Hill.
- Costin, F. (1969). The scrambled sentence test: A group measure of hostility. *Educational and Psychological Measurement*, *29*, 461–468.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.
- Dodge, K. A., & Somberg, D. R. (1987). Hostile attributional biases among aggressive boys are exacerbated under conditions of threat to the self. *Child Development*, *58*, 213–224.
- Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, *22*, 22–37.
- Ellemers, N., Van Rijswijk, W., Roefs, M., & Simons, C. (1997). Bias in intergroup perceptions: Balancing group identity with social reality. *Personality and Social Psychology Bulletin*, *23*, 186–198.
- Gordijn, E. H., Koomen, W., & Stapel, D. A. (2001). Level of prejudice in relation to knowledge of cultural stereotypes. *Journal of Experimental Social Psychology*, *37*, 150–157.
- Herr, P. M. (1986). Consequences of priming: Judgment and behaviour. *Journal of Personality and Social Psychology*, *51*, 1106–1115.
- Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575–604.
- Hewstone, M., & Ward, C. (1985). Ethnocentrism and causal attribution in South East Asia. *Journal of Personality and Social Psychology*, *48*, 614–623.
- Higgins, E. T. (1996). The 'self digest': Self-knowledge serving self-regulatory functions. *Journal of Personality and Social Psychology*, *71*, 1062–1083.
- Islam, M. R., & Hewstone, M. (1993). Intergroup attributions and affective consequences in majority and minority groups. *Journal of Personality and Social Psychology*, *64*, 936–950.
- Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*, 284–308.
- Mummendey, A., Linneweber, V., & Löschper, G. (1984). Aggression: From act to interaction. In A. Mummendey (Ed.), *Social psychology of aggression: From individual behavior to social interaction* (pp. 69–106). New York, NY: Springer.
- Mummendey, A., & Otten, S. (1989). Perspective-specific differences in the segmentation and evaluation of aggressive interaction sequences. *European Journal of Social Psychology*, *19*, 23–40.
- Oldenhuis, H., Gordijn, E. H., & Otten, S. (2005). The Dutch and ethnic minorities in the Netherlands: Stereotypes and meta-stereotypes. Unpublished data. University of Groningen, 2005.
- Otten, S. (2002). 'Me' and 'us' or 'us' and 'them'? The self as a heuristic for defining minimal ingroups. In W. Stroebe, & M. Hewstone (Eds.), *European review of social psychology* (Vol. 13, pp. 1–34). Hove: Psychology Press.
- Otten, S., & Moskowitz, G. B. (2000). Evidence for implicit evaluative in-group bias: Affect-biased spontaneous trait inference in a Minimal Group Paradigm. *Journal of Experimental Social Psychology*, *36*, 77–89.
- Otten, S., & Wentura, D. (1999). About the impact of automaticity in the minimal group paradigm: Evidence from affective priming tasks. *European Journal of Social Psychology*, *29*, 1049–1071.
- Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality & Social Psychology Bulletin*, *5*, 461–476.
- Sachdev, I., & Bourhis, R. Y. (1987). Status differentials and intergroup behaviour. *European Journal of Social Psychology*, *17*, 277–293.
- Sachdev, I., & Bourhis, R. Y. (1991). Power and status differentials in minority and majority group relations. *European Journal of Social Psychology*, *21*, 1–24.
- Spears, R., Gordijn, E., Dijksterhuis, A., & Stapel, D. A. (2004). Reaction in action: Intergroup contrast in automatic behavior. *Personality and Social Psychology Bulletin*, *30*, 605–616.

- 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.
- Stapel, D. A., & Koomen, W. (1998). When stereotype activation results in (counter) stereotypical judgments: Priming stereotype-relevant traits and exemplars. *Journal of Experimental Social Psychology*, *34*, 136–163.
- Stapel, D. A., Koomen, W., & Van der Pligt, J. (1997). Categories of category accessibility: The impact of trait versus exemplar priming on person judgments. *Journal of Experimental Social Psychology*, *33*, 44–76.
- Swim, J. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, *66*, 21–36.
- 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.
- Wigboldus, D. H. J., Dijksterhuis, A., & van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*, 470–484.
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*, 815–827.