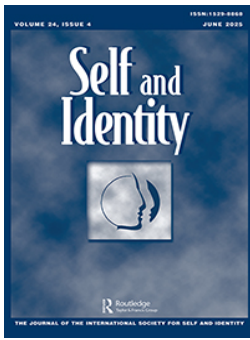


# ULRR

## **Indirect measures of social identities: implicit association tests assess self-categorization, match–mismatch paradigms distinguish identification from disidentification**

Item Type	Article
Authors	Matschke, Christina;Roth, Jenny;Deutsch, Roland;Heineck, Paul-Michael;de Vreeze, Jort
Citation	Self and Identity
Publisher	Routledge Taylor & Francis Group
Download date	2026-04-15 10:21:16
Item License	<a href="https://creativecommons.org/licenses/by-nc-sa/4.0/">https://creativecommons.org/licenses/by-nc-sa/4.0/</a>
Link to Item	<a href="https://doi.org/10.34961/researchrepository-ul.29062967">https://doi.org/10.34961/researchrepository-ul.29062967</a>



## Indirect measures of social identities: implicit association tests assess self-categorization, match–mismatch paradigms distinguish identification from disidentification

Christina Matschke, Jenny Roth, Roland Deutsch, Paul-Michael Heineck & Jort de Vreeze

To cite this article: Christina Matschke, Jenny Roth, Roland Deutsch, Paul-Michael Heineck & Jort de Vreeze (09 May 2025): Indirect measures of social identities: implicit association tests assess self-categorization, match–mismatch paradigms distinguish identification from disidentification, *Self and Identity*, DOI: [10.1080/15298868.2025.2499138](https://doi.org/10.1080/15298868.2025.2499138)

To link to this article: <https://doi.org/10.1080/15298868.2025.2499138>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 09 May 2025.



[Submit your article to this journal](#)



Article views: 20



[View related articles](#)



[View Crossmark data](#)

## Indirect measures of social identities: implicit association tests assess self-categorization, match–mismatch paradigms distinguish identification from disidentification

Christina Matschke<sup>a</sup>, Jenny Roth<sup>b</sup>, Roland Deutsch<sup>c</sup>, Paul-Michael Heineck<sup>c</sup> and Jort de Vreeze<sup>a</sup>

<sup>a</sup>Leibniz-Institut fuer Wissensmedien, Knowledge Construction Lab, Tuebingen, Germany; <sup>b</sup>Department of Psychology, University of Limerick, Limerick, Ireland; <sup>c</sup>Department of Psychology, University of Wuerzburg, Wuerzburg, Germany

### ABSTRACT

A social identity refers to (1) whether people consider a group to be an ingroup along with (2) the psychological meaning of that ingroup (Tajfel, 1978). Social identification and disidentification are two constructs that describe different natures of the psychological meaning of an ingroup. Besides self-reported measures, the Identity Implicit Association Test (identity IAT), and the Match-Mismatch Paradigm (MMP) are the most frequently used measures to assess social identity. In three studies ( $N = 87$ ,  $N = 96$ ,  $N = 137$ ) we tested whether the MMP and identity IAT distinguish between social identification, non-identification, disidentification, and non-categorization. The findings indicate that the identity IAT mostly assesses self-categorization whereas the MMP is sensitive to the specific psychological meaning of an ingroup.

### ARTICLE HISTORY

Received 3 September 2024  
Revised 17 April 2025  
Accepted 23 April 2025

### KEYWORDS


Social identity; indirect measures; disidentification; social identification; implicit identification

Indirect measures became popular in the 1990s, have been critically discussed ever since (Gawronski, 2019; Machery, 2022; Schimmack, 2021; Van Dessel et al., 2020), and gained an established position among the methods to capture psychological constructs. In comparison to direct self-report measures, indirect measures are usually less affected by motivational distortions such as self-presentation, social desirability biases, or demand characteristics (e.g., Gawronski et al., 2020). Mostly implemented to assess attitudes or stereotypes, indirect measures are also used to assess people's social identity (Greenwald & Farnham, 2000).

Social Identity Theory (Tajfel, 1978) defines *social identity* as “that part of the individual's self-concept which derives from his knowledge of his membership of a social group (or groups) together with the emotional significance attached to that membership” (Tajfel, 1978, p. 63). The first aspect of this definition refers to whether people acknowledge a certain group membership and consider a social group to be

**CONTACT** Christina Matschke  [c.matschke@iwm-tuebingen.de](mailto:c.matschke@iwm-tuebingen.de)  Leibniz-Institut fuer Wissensmedien, Knowledge Construction Lab, Schleichstrasse 6, Tübingen 72076, Germany

The first two authors listed equally contributed to the article and share first authorship.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15298868.2025.2499138>

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

an ingroup, thus self-categorizing. The second part of this definition refers to differences in the psychological meaning of an ingroup which can be reflected in the constructs of social identification (Leach et al., 2008; Postmes et al., 2012) and disidentification (Becker & Tausch, 2014; Matschke & Sassenberg, 2010a, 2010b). Social identification and disidentification differ inherently in their nature. Whereas social identification refers to the extent to which people are satisfied and committed to their ingroup (e.g., Leach et al., 2008; Postmes et al., 2012), disidentification refers to the extent to which people devalue their ingroup and distance themselves from it (Becker & Tausch, 2014).

The most frequently used indirect measures to assess social identity are the identity Implicit Association Test (identity IAT, Greenwald et al., 2002) and the Match-Mismatch Paradigm (MMP, Aron et al., 1991). Despite their frequent use, it has never been tested which of the two specific aspects of social identities, i.e., self-categorization or the meaning of an ingroup, they assess. Consequently, it remains unclear whether they distinguish between social identification and disidentification. In the present paper, we test whether an identity IAT and the MMP assess differences in the psychological meaning of an ingroup and are thus suitable to distinguish between social identification and disidentification.

## Psychological meaning of an ingroup: social identification and disidentification

Social identities develop when individuals (1) acknowledge a certain group as an ingroup, thus self-categorizing, and (2) when they attach meaning to that group membership (Tajfel, 1978). The meaning is usually reflected in the amount of *social identification* with the ingroup (Leach et al., 2008; Postmes et al., 2012). However, recent research suggests that individuals may attach qualitatively different meanings to different ingroups (Kreiner & Ashforth, 2004). Whereas they show social identification for some ingroups, they show disidentification from other ingroups.<sup>1</sup> Disidentification is not the opposite of social identification, as that would be indifference (i.e., non-identification). Instead, it is a negative internalized relation to an ingroup that ranges from indifference to strong disidentification. Research has shown that disidentification is, for instance, likely for groups that are low in status (Ikegami & Ischida, 2007), have a negative reputation (Elsbach & Bhattacharya, 2001; Kreiner & Ashforth, 2004), reject the group member (Matschke & Sassenberg, 2010b; Verkuyten & Yildiz, 2007), or provide a low fit between a member and the group (de Vreeze & Matschke, 2019; de Vreeze et al., 2018; Hamstra et al., 2015; Matschke et al., 2023).

Disidentification is defined as a relation to a group where individuals self-categorize as ingroup members, and their self-descriptions, emotions, and behaviors are affected by the ingroup, just like identified peoples', but the psychological meaning of the ingroup is different. Cognitively, identified group members describe themselves as similar to average group members, whereas disidentified members describe themselves as contrary to the ingroup prototype (Becker & Tausch, 2014). Thus, the perception of overlap versus opposition between the description of characteristics of the self and the ingroup is a central defining distinction between social identification and disidentification. Regarding affect, identified group members are

happy to belong to the group, whereas disidentified members are unhappy to belong to the group. Concerning behavior, identified group members are committed to the group's interests and act on its behalf, whereas disidentified people feel detached from the ingroup, distance themselves, and act in contrast to the group's interests.

## Measures of social identities

By acknowledging disidentification as one potential form of social identities, the equation social identity = social identification has been challenged. Most measures of social identities have been developed long before the different meanings of ingroups were established in social psychology. Since self-report measures of social identification usually only differentiated social identification from indifference, but not necessarily indifference from disidentification, scales were developed that are sensitive to capture disidentification (Becker & Tausch, 2014; Matschke & Sassenberg, 2010b). Since then, research that was keen on assessing both dimensions has used both social identification and disidentification scales to grasp the full range of possible psychological ingroup meanings (e.g., Becker et al., 2011; Gligorić & Obradović, 2024; Matschke et al., 2023).

Using both social identification and disidentification scales has proven useful because it reveals relations between social identities and other constructs that might otherwise be undetected. For example, disidentification is especially sensitive to negative group-related experiences, e.g., ingroup rejection (Matschke & Sassenberg, 2010b), low ingroup status (Ikegami, 2010; Kreiner & Ashforth, 2004), or value conflict (De Vreeze & Matschke, 2017, 2019; Elsbach & Bhattacharya, 2001) and is more predictive for *negative* ingroup-related outcomes such as ingroup derogation (Ikegami & Ischida, 2007), anti-normative actions (Matschke & Sassenberg, 2010a; Zou et al., 2008) or counter-group actions and emotions (Becker & Tausch, 2014; De Vreeze & Matschke, 2017; 2019; Elsbach & Bhattacharya, 2001) than low levels of social identification (Ashforth et al., 2013; Becker & Tausch, 2014; De Vreeze & Matschke, 2017; 2019; Kreiner & Ashforth, 2004; Matschke & Sassenberg, 2010a; Matschke et al., 2023). Distinguishing between these different meanings of the ingroup and using measures that capture this distinction thus allows a deeper understanding of self-ingroup relations that affect individuals' cognition, emotions, and behavior.

While direct measures that differentiate social identification and disidentification were developed, indirect measures that specifically assess disidentification are, so far, still missing. Indirect measures have the advantage to uncover social identities when social desirability or self-presentation renders explicit expressions of social identities difficult, such as when group memberships are undesirable or stigmatized, or when group members are unaware of their social identities (Coats et al., 2000).

The most frequently used indirect measures of social identity, the identity IAT (Greenwald et al., 2002) and the MMP (Aron et al., 1991), have, to date, not yet been tested whether they differentiate between social identification and disidentification. Correlations between indirect and self-reported measures of social identities provide insight into the overlap of two measurement approaches. However, those correlations appear to depend on the implemented measures and are either insignificant (Wang et al., 2022) or weak to moderate in size (e.g., Coats et al., 2000; Sassenberg & Matschke, 2010).

Thus, there is reason to believe that there is considerable divergence between indirect and self-reported measures of social identity (Theodoridis, 2017).

Indirect measures of social identities are considered an important tool to better understand people's social identities, and identity IATs and the MMP have gained an established position amongst measures of social identities. To advance theorizing and draw conclusions for their application, it is, however, essential to understand what the indirect scores represent and what its limits are. We argue that by increasing our understanding of what these indirect measures of social identities do (and do not) assess, they can be used to be more targeted and contribute to more elaborate theory development and application.

### The use and interpretation of indirect measures of social identities

Identity IATs are widespread and have been implemented in a variety of studies where the self-ingroup relation is explored and used to predict other constructs. Identity IATs are central to Balanced Identity Theory (Cvencek et al., 2012, 2021; Greenwald et al., 2002) that was developed to predict behavior (Hughes & Smith, 2025), intergroup relations (Loughnane et al., 2023), and social identity change (Loughnane et al., 2021; Roth et al., 2018).

Typical areas of use are stigmatized identities (e.g., Grover et al., 2003; Kirby et al., 2021), professional identities (e.g., Wang et al., 2022), and political identities (e.g., Hawkins & Nosek, 2012; Theodoridis, 2017). Another area of application of variants of identity IATs is research on drugs (e.g., Chen et al., 2021) and drinking (e.g., Albery et al., 2024; Gray et al., 2011) where researchers aim to assess people's relationships with social groups and respective objects by circumventing self-presentation concerns.

However, the interpretation of identity IAT is still subject to discussion in the literature. While some researchers interpret identity IAT scores as indicators of social identification (e.g., Aidman & Carroll, 2003; Chen et al., 2021; Cvencek et al., 2011; Devos & Torres, 2007; Hawkins & Nosek, 2012; Kirby et al., 2021; Rudman & Goodwin, 2004), other researchers label it social identity (Cvencek et al., 2016, 2021), a measure of identity centrality (Clifton et al., 2021), and yet others criticize its interpretation as a measure of social identification and suggest it to be a measure of self-categorization (Cadinu & Galdi, 2012).

The MMP is another indirect measure of social identity. It is usually used to differentiate ingroups from outgroups (Coats et al., 2000; Smith & Henry, 1996; Smith et al., 1997), to detect the formation of social identities for novel ingroups after (vs. without) contact (Sassenberg & Matschke, 2010), or to differentiate groups that are relevant (vs. irrelevant) for the self-concept (Tropp & Wright, 2001). Usually, the literature does not explicitly label the MMP score as a measure of social identification but as a measure of social identities (Smith & Henry, 1996; Smith et al., 1997). Social identities, however, might theoretically include both social identification and disidentification, and the score is usually interpreted in the sense of social identification (e.g., Coats et al., 2000; Sassenberg & Matschke, 2010; Smith & Henry, 1996; Tropp & Wright, 2001). Coats et al. (2000) explicitly interpret their finding that the match-mismatch effect from the MMP correlates with direct measures of social identification as evidence for the construct validity of the match-mismatch effect as a measure of social identification. Similarly, the finding that the match-mismatch effects for ingroups are stronger for high compared to low identifiers (Tropp & Wright, 2001), and

for people with identification motives vs. differentiation motives, Brewer & Pickett (1999) was interpreted as evidence for the validity that the effect captures social identification.

Taken together, both indirect measures of social identities are often interpreted as differentiating high and low identifiers. It remains unclear, however, how disidentified group members figure in identity IATs and MMPs and whether these measures distinguish between social identification and disidentification at all. The lack of knowledge about which the specific aspect of social identities (i.e., self-categorization or meaning of an ingroup) is captured by identity IATs and the MMP, and which results they produce if the meaning of an ingroup is negative may, however, easily lead to incomplete or wrong interpretations of the results. In the following, we elaborate on the procedure of both indirect measures and argue that they are prone to measure different aspects of social identity.

### The identity implicit association test

Identity IATs aim to assess the relative strength of the mental connection between the self (compared to other) and an ingroup relative to the mental connection between the self (compared to other) and an outgroup (Greenwald et al., 2002). Participants' task is to categorize labels for self (or Me; e.g., me, mine, I) and other (or not-me; e.g., they, their, them) and labels of the groups of interest (e.g., female vs. male) using two response keys. In a gender identity IAT, for example (Aidman & Carroll, 2003; Nosek et al., 2002; Rudman & Goodwin, 2004), people in one block categorize labels for Self and Female with one computer key as well as Other and Male with a different key. In another block, they categorize labels for Self and Male with one key and Other and Female with another key. In each of these tasks, response latencies are assessed. Most of the research that implemented an identity IAT assumes that people who identify more strongly with the ingroup show faster responses when the ingroup and the self share the same response key (compatible trials) than when the outgroup and the self share the key (incompatible trials; Aidman & Carroll, 2003; Cvencek et al., 2016; Nosek et al., 2002; Rudman & Goodwin, 2004). Consequently, if reaction times in the compatible trials are faster than the incompatible trials, it is derived that participants have a stronger mental connection between the self and that group than with the other group. Typically, this pattern of results is interpreted as implicit social identification (e.g., Aidman & Carroll, 2003; Devos et al., 2010; Theodoridis, 2013, 2017), implicit social identity (Cummins et al., 2021; Cvencek et al., 2014, 2016; Devos, 2006), identity centrality (Clifton et al., 2021), or self-categorization (Cadinu & Galdi, 2012).

We argue that an identity IAT captures whether people subjectively belong to one group or the other (i.e., self-categorization) but may not be sensitive to assess differences in the psychological meaning of the ingroup. Identity IATs include ingroup/outgroup and self/other labels without reference to category-content or the psychological meaning of the groups. Thus, identity IATs assess primarily the strength of the mental connection between labels of self and group. A strong mental connection between self and group labels primarily indicates that the group is part of their self-concept. We assume, however, that identity IATs are not able to differentiate between social identification and disidentification because the main differences between social identification and disidentification on the cognitive (perception of attribute overlap vs. opposition), affective (positive vs.

negative affect), and behavioral (behavior in line vs. in contrast to the group's interests) dimension are not content of the IAT measurement. Notably, disidentified people – like identified people – usually consider the ingroup to be a relevant part of the self-concept. As such, they may show a similar strong identity IAT effect.

Why should disidentified group members show a mental self-ingroup connection of comparable strength as identified group members? Research indicates that mental connections between self and group can, for instance, result from repeated co-activation (e.g., Hutchison, 2003; Lucas, 2000, for reviews). Co-activation can occur for any pair of concepts that are often activated together, be it by simultaneous presentation of the two concepts (i.e., phrasal associates, e.g., bread and butter, or functional associates, e.g., hammer and nail), or the belonging to the same category (e.g., mother and father). As such, even supposedly contrary concepts (i.e., antonyms, e.g., dark and bright) and concepts that are coupled by a negation can be mentally connected (Hutchison, 2003; Lucas, 2000; Mayo et al., 2004). Based on this evidence, we argue that disidentified group members show a self-ingroup connection because, for instance, the concepts of the self and the ingroup are activated simultaneously many times. For example, some group memberships are given (e.g., nationality or physical stigma), or members are forced into groups by circumstances (e.g., social class or employment status, e.g., Barreto & Ellemers, 2003; Deaux, 1996). Even if not self-selected or valued, the group member must be in touch with that group and will be perceived and treated as a member by others. Furthermore, actively distancing oneself from an ingroup and defining oneself in opposition to the group prototype (i.e., frequent co-activation of the self and the ingroup because the concepts are coupled with negation) can lead to a strong self-group connection. In addition, recent research has demonstrated that inferences and reasoning impact results on indirect measures (Charlesworth et al., 2020; de Houwer, 2009; Kurdi & Banaji, 2019; Mitchell et al., 2009). Therefore, if people hold the conviction that they factually belong to the ingroup (even if they do not like it), this self-categorizing knowledge can be expected to affect identity IATs. In summary, we therefore predict that identity IATs will not differentiate between ingroup social identification and disidentification.

### The match-mismatch paradigm

Just like identity IATs, MMPs do not assess the psychological meaning of an ingroup directly but infer this from the individual's performance in a decision task under time pressure. The difference between identity IATs and the MMP is that an identity IAT captures whether labels of self and group are mentally connected, whereas the MMP compares trait-based matches (overlaps) versus mismatches (oppositions) between self and group.

The MMP (Aron et al., 1991) is based on the rationale that when individuals identify with a social group, the mental representations of the self and the group are strongly connected and activate each other automatically (Smith, 2002; Smith et al., 1999). In the MMP, participants first rate themselves and two groups (an ingroup and a control outgroup) on several traits. From these descriptions, matches (traits that self and group share) and mismatches (unshared traits) for the ingroup and outgroup are identified for every participant. Irrelevant traits (where group or self are rated indifferently) are not considered. Afterwards, the same traits are presented one by one, and

participants decide as quickly as possible whether the respective trait describes them or not, while response latencies are assessed. Smith and colleagues claim that if the self-group connection leads to a mutual activation of the mental representations, matches receive activation from two sources (by the connection of the trait to both self and group), which will result in faster response times than mismatches (Smith, 2002; Smith et al., 1999). The average difference between reaction times to matching and mismatching traits for the ingroup is therefore calculated (the match-mismatch-effect): faster matches than mismatches with an ingroup are interpreted as a sign of the existence of a social identity.

Other than the IAT, by assessing the acceptance/rejection of traits, the MMP captures descriptive content of the concepts self and group and indirectly allows participants to position themselves in line with or contrary to the group characteristics, thus tapping into the cognitive difference between social identification and disidentification. In the MMP, reaction times of matches (compared to mismatches) with the ingroup should be facilitated for identified people, because matches are channeled in memory: Describing oneself in line with the ingroup's traits and rejecting contrary traits is the usual behavior of identified members, whereas describing oneself opposing to the ingroup is not. For disidentified group members, rejecting typical ingroup characteristics and accepting opposing characteristics should be channeled in memory (Zou et al., 2008), which should facilitate mismatches in reaction times and reduce the match-mismatch effect. We therefore expect that the MMP differentiates disidentification and social identification.

## Overview and hypotheses

Research conducting the MMP and identity IATs often interprets the indirect scores as measures of social identification. We argue that this might lead to a false conclusion because a) it uses the degree of psychological meaning of an ingroup and self-categorization interchangeably and b) it does not take different qualities of meanings of ingroup (i.e., social identification and disidentification) into account.

The present research therefore tests what aspect the two established indirect measures of social identity capture and whether they differentiate between social identification and disidentification. Three studies<sup>2</sup> manipulate the meaning of the group and test its effect on the identity IAT and the MMP. We add a non-identification condition for explorative purposes because non-identification theoretically is the zero point of social identification and disidentification. This results in three levels of the meaning of the group (disidentification vs. non-identification vs. social identification). In studies 2 and 3, we add a non-categorization condition where people are not members of any of the social groups investigated. Hierarchical linear modeling is used to test the predictions across all data sets. We hypothesized that both the MMP and the identity IAT differentiate between the ingroup conditions and the non-categorization condition, while only the MMP differentiates between disidentification and identification with the ingroup. In the MMP, we expect that group members who identify with the ingroup differ from those who disidentify in their ingroup match-mismatch effect.

We preregistered Study 3 before data collection ([https://aspredicted.org/LYB\\_3T9](https://aspredicted.org/LYB_3T9)). We report all measures, manipulations, and exclusions in the respective method sections. The

ethics committee of the Leibniz-Institut für Wissensmedien (Tübingen) has approved the study series (LEK 2013/002 and LEK 2023/011).

## Study 1

Study 1 tested whether the MMP and an identity IAT distinguish between social identification and disidentification.

### Method

#### Design

Study 1 had a one-factorial between-subjects design manipulating the meaning of the ingroup with three conditions: disidentification condition (DIDC) vs. social identification condition (SIDC) vs. non-identification condition (NIDC). The identity IAT-D effect and the ingroup and outgroup match-mismatch effects in the MMP served as dependent variables.

#### Participants

At a German university, a convenience sample of 87 undergraduates that could be recruited at that time for the laboratory study participated in exchange for 4€. No stopping rule was used, and the sample size depended on how many participants we were able to recruit for the period of lab time booked. We add sensitivity analysis for our central hypotheses' tests in the results section.

In line with the exclusion criteria (see below), we excluded participants who did not follow the instructions ( $n = 6$ ) and whose MMP data were incomplete ( $n = 8$ ) or outliers ( $n = 1$ ). The analyzed sample includes  $N = 72$  (49 females, 23 males,  $M_{Age} = 26$ ,  $SD_{Age} = 7.94$ , range: 19–62 years). We assigned participants randomly to the DIDC ( $n = 27$ ), the SIDC ( $n = 28$ ), or the NIDC ( $n = 17$ ).

#### Procedure

Participants first indicated a social group in line with the instructions that manipulated the meaning of the group. Then, participants filled in two items on the evaluation and emotional bond to the ingroup<sup>3</sup> and completed the identity IAT and the MMP. The order of the IAT and MMP was counterbalanced. Then, participants completed self-report measures of social identification and disidentification to validate the manipulation and to assess whether the self-report measures differentiate between the constructs of social identification and disidentification. Finally, we included exploratory items and items to validate the findings,<sup>4</sup> assessed demographics, and debriefed, thanked, and compensated participants.

#### Manipulation of the meaning of the ingroup

We manipulated the meaning of the group through the following instruction:

We all belong to different groups that are of different meaning to us. This study is about groups that you belong to. Think of groups with whose members you share attributes or of groups where a common function or a common goal is the focus of the group. This

study is not about groups where only personal ties between single members are in the focus, but about groups that would further exist even if single members would leave the group.

*Disidentification:* This is about a group to which you do **not like to** belong. Please think about which of the groups you belong to, you do not appreciate, you do not fit into and from which you would like to distance yourself, even though you are a member.

*Social identification:* This is about a group you **very much like** to belong and to which you strive to be a member of the group. Please think about which of the groups you belong to, you do appreciate a lot, you do really fit into well and to which you feel a strong sense of belonging to as a member.

*Non-identification:* This is about a group to which you belong to, but where this does **not mean anything** to you. Please think about which of the groups you belong to, you do not at all find relevant, you do neither appreciate nor suitable, nor unsuitable, even though you are a member.

We asked participants to indicate a group that fulfills the respective criteria. To ease this choice, participants saw examples (e.g., religious groups, political parties, and nationality) that other participants had named in pretests of the instructions.

## Measures

**Self-reported social identification and disidentification.** We assessed *social identification* with 15 items<sup>5</sup> (e.g., “I feel committed to [ingroup],”  $\alpha = .95$ , Leach et al., 2008; Roth & Mazziotta, 2015). We measured *disidentification* with 11 items (Becker & Tausch, 2014; e.g., “I’m the opposite of people in this group [group],”  $\alpha = .92$ , original German items). All items are assessed on a 7-point scale (1 = *does not apply at all*, 7 = *applies completely*). Items were presented in a fixed random order, and means served as scale values, with higher means indicating stronger social identification or disidentification, respectively.

**Identity IAT.** The identity IAT followed the procedure by Greenwald et al. (2002). Since participants selected their group individually, they first typed in three words descriptive for their ingroup to gain the exemplars for the classification task. To help them name exemplars, we presented examples.

The IAT consisted of five blocks, three practice blocks, and two combined blocks. Participants started the IAT with two practice blocks (12 trials each). First, they classified three stimuli to the concept category “self” (I, mine, self) and three stimuli to the concept category “other” (those, whose, other). In the second practice block, they classified the three stimuli that they had named for their ingroup. Since the IAT is a relative measure, we used the outgroup “Japanese” as a reference category in all conditions. Therefore, participants also practiced categorizing three stimuli into the category “Japanese” (male version of Japanese, female version of Japanese, Japan). In the first combined block (self/group, other/Japanese) participants classified stimuli of the self and ingroup category on the same computer key (a) and stimuli of the other and Japanese category with a different key (numpad 5). Then, key assignments of the groups were switched, and participants practiced the reversed categorization for another 12 trials. Finally, participants performed the second combined block (self/Japanese, other/ingroup) in which the group categories were reversed. The two

combined blocks (48 trials each) were each separated by a short break after the first 24 trials. The combined block order was held constant since we were mainly interested in differences between experimental conditions (Egloff & Schmukle, 2002). Stimuli are presented in a random order with an inter-trial interval of 150 milliseconds. Category labels were displayed in the left and right upper corners of the screen. A red X that remained on the screen until the correct response was made denotes incorrect responses. In each trial, the reaction time of the first response was stored.

We computed the IAT-D effect following the improved scoring algorithm (Spearman Brown coefficient for test halves = .54; Greenwald et al., 2003). More positive IAT-D effects indicated stronger connections between self and ingroup (and other and Japanese) than between self and Japanese (and other and ingroup).

**MMP.** The MMP was conducted similarly to the paradigm by Smith and Henry (1996). First, participants rated 91 traits (e.g., intelligent, aggressive; Sassenberg & Matschke, 2010) on a 7-point scale (1 = *extremely unlike*, 7 = *extremely like*) for the ingroup, for themselves, and for the outgroup of Japanese. After 10 trial answers, the same traits appeared in random order in the middle of the screen, preceded by a fixation cross of 500 milliseconds. Participants decided as quickly and accurately as possible whether the trait described them or not (yes vs. no) and reaction times were assessed. After the decision, the screen remained blank for 750 milliseconds before the next trait appeared.

We computed the ingroup and outgroup match–mismatch index from the response latencies (Coats et al., 2000; Sassenberg & Matschke, 2010). Ratings for the groups and the self were dichotomized (1–3 = *no*, 5–7 = *yes*). The reaction-time-based group match-mismatch effect was computed by subtracting the mean response times to matches from the mean response times to mismatches. Higher values of the match-mismatch effect are commonly interpreted as stronger inclusion of the group into the self-concept.<sup>6</sup> Table 3 presents correlations between self-reported and indirect measures.

### Exclusion criteria

We applied the following exclusion criteria in all experiments: participants (1) who do not follow instructions (i.e., who indicate more than one social group and those who are not a member of the indicated group), (2) who show less than 10 ratings above or below 4 (i.e., indifference) in the trait assessment of ingroup, outgroup, or self, as part of the MMP (3) who have missing values for either matches or mismatches for the ingroup or outgroup (i.e., match-mismatch indices cannot be calculated), (4) who are outlier in the MMP index ( $>3 SD$ , Sassenberg & Matschke, 2010), and (5) who have more than 10% reaction times below 300 milliseconds in the IAT (Greenwald et al., 2003). In addition, for the calculation of the individual match-mismatch index, reaction times below 300 and above 5000 milliseconds are excluded from the calculation (Smith & Henry, 1996).

### Results

All predictions were tested with a significance level of  $\alpha = .05$  (two-tailed). Data were analyzed with IBM SPSS Statistics Version 29.0 (IBM Corp, 2023). Confidence intervals for effect sizes were computed using R (Version 4.0.5; R Core Team, 2021) and the MBESS package (Version 4.9.3; Kelley, 2023). We report 95% CI for all effect sizes that can be

positive and negative (Cohen's  $d$ ). However, in line with the recommendations of Lakens (2014), we report 90% CI for effect sizes that can only be positive ( $\eta^2$ ), because the 90% CIs around  $\eta^2$  equals the 95% CIs for Cohen's  $d$  effect sizes. While a non-significant CI for Cohen's  $d$  would include 0, a non-significant CI for  $\eta^2$  would start from 0 because  $F$ -tests are always one-sided.

### Self-reported social identification and disidentification

A MANOVA with Meaning of the Group as factor and social identification and disidentification as dependent variables showed significant differences between the conditions,  $F(4,138) = 17.55$ ,  $p < .001$ ,  $\eta_p^2 = .34$ , 90% CI [.22, .41] (Table 1). The conditions differed in social identification,  $F(2,69) = 34.62$ ,  $p < .001$ ,  $\eta_p^2 = .50$ , 90% CI [.35, .59]. Post-hoc tests demonstrate that social identification was higher in the SIDC than in the NIDC,  $t(43) = 5.69$ ,  $p < .001$ ,  $d = 1.75$ , 95% CI [1.04, 2.45], and in the DIDC,  $t(53) = 7.93$ ,  $p < .001$ ,  $d = 2.14$ , 95% CI [1.47, 2.80], whereas the self-reported social identification measure did not differentiate between the NIDC and the DIDC,  $p = .643$ . The conditions also differed in self-reported disidentification,  $F(2,69) = 29.12$ ,  $p < .001$ ,  $\eta_p^2 = .46$ , 90% CI [.30, .56]. Disidentification in the DIDC was higher than in the NIDC,  $t(42) = 4.10$ ,  $p < .001$ ,  $d = 1.27$ , 95% CI [0.60, 1.93], and in the SIDC,  $t(53) = 7.59$ ,  $p < .001$ ,  $d = 2.05$ , 95% CI [1.38, 2.70]. Additionally, disidentification was higher in the NIDC than in the SIDC,  $t(43) = 2.53$ ,  $p = .041$ ,  $d = 0.78$ , 95% CI [0.15, 1.40]. Thus, the manipulation was successful. Moreover, the findings add to the body of evidence that both scales reliably differentiate the respective construct from non-identification, but the social identification scale does not capture the difference between non-identification and disidentification.

**Table 1.** Descriptive statistics of the self-reported and indirect measures based on experimental conditions in study 1 ( $N = 72$ ).

Condition	Dis-identification		Social Identification		Non-Identification	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Social ID	2.97 <sub>a</sub>	1.24	5.48 <sub>b</sub>	0.86	3.42 <sub>a</sub>	1.48
DisID	3.58 <sub>a</sub>	1.21	1.54 <sub>b</sub>	0.67	2.32 <sub>c</sub>	1.05
<i>MM effect</i>						
Ingroup	36.23 <sub>a</sub>	20.71	133.42 <sub>b</sub>	20.33	45.58 <sub>a</sub>	26.10
Outgroup	60.80 <sub>a</sub>	18.54	75.69 <sub>a</sub>	18.21	70.21 <sub>a</sub>	23.37
<i>IAT-D effect</i>	0.69 <sub>a</sub>	0.40	0.74 <sub>a</sub>	0.39	0.74 <sub>a</sub>	0.44

MM effect = Match-Mismatch Effect Ingroup; ID = Identification; DisID = Dis identification. Different subscripts within the line indicate statistically significant differences between conditions.

### Identity IAT effect

An ANOVA including the IAT effect as dependent variable and Meaning of the Group as independent variable along with Bonferroni corrected post-hoc comparisons tested for differences between conditions. In line with the hypothesis, the analyses showed that the IAT effect (see Table 1) did not differ between SIDC, DIDC, and NIDC,  $F(2,69) = 0.13$ ,  $p = .879$ ,  $\eta_p^2 < .01$ , 90% CI [.00, .03]. A sensitivity analysis (Faul et al., 2007) for this ANOVA with  $\alpha = .05$ , Power = .80,  $N = 72$ , and three between-subjects conditions showed a necessary effect size of  $\eta_p^2 = .125$  to result in a detectable difference between the conditions.<sup>7</sup>

Complementing those findings, *t*-tests against 0 showed a positive IAT effect in each of the conditions, DIDC:  $t(26) = 8.99, p < .001, d = 1.73, 95\% \text{ CI } [1.12, 2.33]$ , NIDC:  $t(16) = 6.92, p < .001, d = 1.68, 95\% \text{ CI } [0.92, 2.42]$ , SIDC:  $t(27) = 10.23, p < .001, d = 1.93, 95\% \text{ CI } [1.29, 2.56]$ , indicating that the identity IATs measured some meaningful difference in responses to the ingroups and outgroup.

### Match–mismatch effects

To test whether the match–mismatch ingroup (but not outgroup) effect differs between social identification and disidentification, we conducted a mixed model ANOVA with the between-group factor Meaning (DIDC vs. SIDC vs. NIDC) and the repeated measure factor Group (i.e., ingroup vs. outgroup). The predicted Meaning  $\times$  Group interaction was significant, Pillai's Trace  $F(2,69) = 3.53, p = .035, \eta_p^2 = .09, 90\% \text{ CI } [.004, .20]$  (Figure 1). The sensitivity analysis (Faul et al., 2007) with  $\alpha = .05, \text{ power} = .80, N = 72$ , three conditions, and a correlation between measures of  $r = .22$ , showed for this interaction a necessary effect size of  $\eta_p^2 = .05$  to reach significance. Additionally, the mixed-model ANOVA showed a significant main effect of meaning,  $F(2,69) = 3.79, p = .027, \eta_p^2 = .10, 90\% \text{ CI } [.006, .20]$ , but not for group,  $F(1,69) = 0.03, p = .855$ .

To decompose the interaction, we conducted three pairwise comparisons (Table 1). First, we compared the ingroup match–mismatch effect between the conditions. As predicted, the ingroup match–mismatch effect was significantly stronger in SIDC than in DIDC,  $t(53) = 3.35, p = .001, d = 0.90, 95\% \text{ CI } [0.34, 1.46]$ . Furthermore, the ingroup match–mismatch effect was also stronger in the SIDC than in the NIDC,  $t(43) = 2.66, p = .010, d = 0.82, 95\% \text{ CI } [0.19, 1.44]$ . The DIDC and NIDC did not differ significantly,  $p = .780$ .

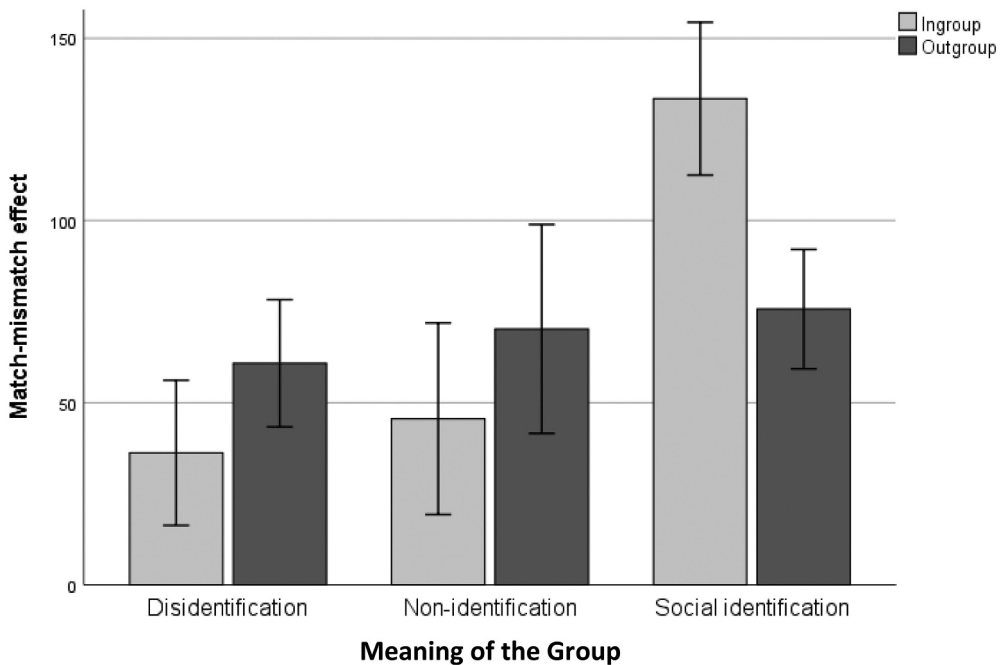


Figure 1. Mean match-mismatch effects (with  $\pm 1$  SE) in Study 1 ( $N = 72$ ).

Second, the outgroup match-mismatch effects were compared between the conditions: there were no significant differences, all  $ps > .56$ . Thus, the results confirmed the hypothesis that the MMP differentiates between ingroup social identification and disidentification.

Third, in the tradition of the MMP, we compared the match-mismatch effects for ingroup and outgroup within each condition. Pairwise comparisons showed that in the SIDC, the match-mismatch effect of the ingroup was stronger than the effect of the outgroup,  $t(27) = 2.38$ ,  $p = .020$ ,  $d = 0.64$ , 95% CI [0.10, 1.17], whereas ingroup and outgroup did not differ significantly in the DIDC, nor in the NIDC, both  $ps > .32$ .

## Discussion

Study 1 indicated that, as predicted, the identity IAT did not differentiate between social identification (SIDC) and disidentification (DIDC). Notably, the present study was only powered to find a medium to strong effect size. Therefore, it is still possible that the IAT differentiates between conditions in a larger sample.

As predicted, the MMP demonstrated stronger ingroup (but not outgroup) match-mismatch effects in the SIDC than DIDC. Comparing the ingroup and outgroup effects within the conditions, only the SIDC showed stronger effects for the ingroup than for the outgroup. The data even point to a slightly reverse (but not significant) pattern in the DIDC, where the match-mismatch effect for the outgroup was descriptively larger than for the ingroup. Thus, the results provide initial evidence that the MMP effect is sensitive to differences in the psychological meaning of an ingroup in that it distinguished between social identification and disidentification, whereas the identity IAT did not. Despite showing a substantially larger mental self-ingroup connection than mental self-outgroup connection as reflected in the IAT effect, there is no strong indication of whether this effect reflects disidentification, non-identification, or social identification.

In line with previous research (Becker & Tausch, 2014), the social identification scale did not differentiate non-identification from disidentification, whereas the disidentification scale did. This finding supports the notion that using scales for both constructs contributes to a more comprehensive picture of the psychological meaning of an ingroup.

Notably, the size of the absolute IAT effect in Study 1 is partly due to the fixed order of IAT tasks and thus, the absolute IAT effect is overestimated in reflecting stronger self-ingroup compared to mental self-outgroup connections (Greenwald et al., 1998). Study 2 therefore counterbalanced the order of the IAT test blocks.

## Study 2

While study 1 provided initial evidence that the MMP captures differences in the psychological meaning of an ingroup, it provided little evidence on what identity IAT is assessing. Cadinu and Galdi (2012) suggested that identity IATs might be particularly sensitive to self-categorizations and considered the identity IAT to capture whether a person belongs to one group or the other. To make sure that the current study setup is generally able to map manipulations relevant to identity IATs, we added a condition where participants belong to neither group in Study 2: the non-categorization condition (NCATC). In all conditions, we maintained the reference outgroup Japanese, but in the NCATC, the group

Koreans was introduced as a second outgroup that replaced the ingroup used in the other conditions. Given that mental self-ingroup connections are stronger than mental self-outgroup connections (which make the identity IAT effect), we predicted a substantially smaller IAT effect in the NCATC than in all ingroup conditions. As in Study 1, we again predicted stronger ingroup (but not outgroup) match-mismatch effects in the SIDC compared to the DIDC.

## Method

### Participants

We used a convenience sample with the maximum number of participants that we could recruit for the lab time booked (without a stopping rule). In exchange for 4€, 110 students participated in the study at a different German University. Based on the same exclusion criteria for Study 1, we excluded  $n = 4$  who did not follow instructions,  $n = 7$  whose data were incomplete in the MMP, and  $n = 3$  outliers in the match-mismatch effects. We analyzed the data of 96 participants (59 females, 37 males,  $M_{\text{Age}} = 25$ ,  $SD_{\text{Age}} = 5.01$ , range: 19–49 years). Participants were randomly allocated to the conditions: DIDC ( $n = 25$ ), NIDC ( $n = 25$ ), SIDC ( $n = 23$ ), and NCATC ( $n = 23$ ).

### Procedure and measures

The procedure of Study 2 was the same as in Study 1, but we counterbalanced the order of the test blocks of the identity IAT to control for order effects that influence the IAT effect (Greenwald et al., 1998). We computed the IAT-D-effect (Spearman Brown coefficient for test halves = .77) and the ingroup and outgroup effects in the MMP and averaged all items of the respective self-reported scales (social identification,  $\alpha = .93$  and disidentification,  $\alpha = .89^8$ ). The presentation of items of social identification and disidentification scales were randomized within the questionnaires. Again, explorative items were added to validate the manipulation check<sup>4</sup>.

In the NCATC condition, participants read:

We are all familiar with various social groups that have different meaning to us. This study is about social groups that you do not belong to. This study is about Koreans and Japanese.

In this condition, all items that refer to the ingroup in the other conditions refer to the group of Koreans.

## Results

### Self-reported social identification and disidentification

The MANOVA with Meaning as independent variable and self-reported social identification and disidentification as dependent variables demonstrated significant differences between the conditions,  $F(6,184) = 14.10$ ,  $p < .001$ ,  $\eta_p^2 = .32$ , 90% CI [.21, .38] (Table 2). The conditions differed in social identification,  $F(3,92) = 20.91$ ,  $p < .001$ ,  $\eta_p^2 = .41$ , 90% CI [.26, .49]. Post-hoc comparisons showed that social identification was higher in the SIDC than in the DIDC,  $t(46) = 5.94$ ,  $p < .001$ ,  $d = 1.72$ , 95% CI [1.04, 2.38] and higher than in the NIDC,  $t(46) = 4.67$ ,  $p < .001$ ,  $d = 1.35$ , 95% CI [0.71, 1.97], and the NCATC,  $t(46) = 7.53$ ,  $p < .001$ ,  $d = 2.22$ , 95% CI [1.47, 2.95]. As in study 1, social identification between DIDC and NIDC or

**Table 2.** Descriptive statistics of the self-reported and indirect measures' indices based on experimental conditions in study 2 ( $N = 96$ ).

	Dis-identification		Social identification		Non-identification		Non-categorization	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Social ID</i>	3.59 <sub>ab</sub>	1.04	5.30 <sub>c</sub>	0.92	3.96 <sub>b</sub>	1.27	3.09 <sub>ad</sub>	0.61
<i>DisID</i>	3.59 <sub>a</sub>	1.13	1.92 <sub>c</sub>	0.92	2.33 <sub>bc</sub>	0.89	2.85 <sub>ab</sub>	0.99
<i>MM effect</i>								
Ingroup	33.45 <sub>ac</sub>	26.88	132.37 <sub>bd</sub>	27.46	63.13 <sub>cd</sub>	26.33	74.47 <sub>cd</sub>	28.07
Outgroup	101.25 <sub>a</sub>	18.42	84.08 <sub>a</sub>	18.82	97.29 <sub>a</sub>	18.05	89.67 <sub>a</sub>	19.24
<i>IAT-D effect</i>	0.43 <sub>a</sub>	0.61	0.55 <sub>a</sub>	0.51	0.59 <sub>a</sub>	0.43	-0.24 <sub>b</sub>	0.39

MM effect = Match-Mismatch Effect Ingroup; ID = Identification; DisID = Dis identification. In the non-categorization condition, ingroup was replaced by the outgroup Koreans. Different subscripts within the line indicate statistically significant differences between conditions.

**Table 3.** Correlations of self-report and indirect measures in studies 1 ( $N = 72$ )/2 ( $N = 96$ )/3 ( $N = 137$ ).

Variable	1	2	3	4
1. IAT-D effect				
2. IG MM effect	-.05/.04/-.03			
3. OG MM effect	.06/-.05/.08	.22 <sup>+</sup> /.31 <sup>**</sup> /.23 <sup>**</sup>		
4. Social ID	.15/.42 <sup>**</sup> /.43 <sup>***</sup>	.34 <sup>**</sup> /.16/.12	.01/-.03/.02	
5. DisID	-.07/-.18 <sup>+</sup> /.19 <sup>*</sup>	-.30 <sup>*</sup> /.19 <sup>+</sup> /.16 <sup>+</sup>	.09/-.03/.01	-.68 <sup>***</sup> /.63 <sup>***</sup> /.66 <sup>***</sup>

IG MM effect = Match-Mismatch Effect Ingroup; OG MM effect = Match-Mismatch Effect Outgroup; ID = Identification; DisID = Dis identification, <sup>+</sup> $p < .10$ ; <sup>\*</sup> $p < .05$ ; <sup>\*\*</sup> $p < .01$ ; <sup>\*\*\*</sup> $p < .001$ .

NCATC did not differ, both  $ps > .50$ , but social identification was stronger in the NIDC than in the NCATC,  $t(46) = 3.02$ ,  $p = .020$ ,  $d = 0.87$ , 95% CI [0.27, 1.46]. The conditions also differed in disidentification,  $F(3,92) = 12.80$ ,  $p < .001$ ,  $\eta_p^2 = .29$ , 90% CI [.15, .39]. In the DIDC, participants showed higher disidentification than participants in the NIDC,  $t(48) = 4.50$ ,  $p < .001$ ,  $d = 1.27$ , 95% CI [0.66, 1.88]. and in the SIDC,  $t(46) = 5.84$ ,  $p < .001$ ,  $d = 1.69$ , 95% CI [1.02, 2.34], but the DIDC did not differ significantly from the NCATC,  $t(46) = 2.61$ ,  $p = .063$ ,  $d = 0.76$ , 95% CI [0.17, 1.34]. The SIDC was significantly lower than the NCATC,  $t(44) = -3.16$ ,  $p = .013$ ,  $d = -0.93$ , 95% CI [-1.54, -0.32], and all other differences were not statistically significant,  $ps > .46$ . Thus, the data confirm the effectiveness of the manipulation and demonstrate the sensitivity of the self-reported measures for the respective construct.

### Identity IAT effect

In all ingroup conditions, positive IAT effects indicate stronger self-ingroup connections than self-Japanese connections. In the NCATC, positive IAT effects indicate stronger connections of self and Koreans than self and Japanese. To test whether the IAT effect is smaller in the NCATC than in the ingroup conditions (i.e., SIDC, DIDC, and NIDC), we conducted an ANOVA with the IAT effect as dependent variable and Meaning as independent variable.

Conditions differed significantly,  $F(3,91) = 14.31$ ,  $p < .001$ ,  $\eta_p^2 = .32$ , 90% CI [.18, .42] (Table 2). Post-hoc comparisons demonstrated that the IAT effect was lower in the NCATC than in the SIDC,  $t(46) = -5.44$ ,  $d = -1.60$ , 95% CI [-2.26, -0.93], the NIDC,  $t(46) = -5.78$ ,  $d = -1.67$ , 95% CI [-2.32, -1.00], and the DIDC,  $t(46) = -4.65$ ,  $d = -1.35$ , 95% CI [-1.97, -0.71], all  $ps < .001$ . Replicating results of study 1, none of the ingroup conditions

differed from each other in the IAT effect, all  $ps = 1.000$ . As in Study 1,  $t$ -tests against 0 showed a positive IAT effect in each of the ingroup conditions, DIDC:  $t(23) = 3.46$ ,  $p = .002$ ,  $d = 0.71$ , 95% CI [0.25, 1.15], NIDC:  $t(24) = 6.80$ ,  $p < .001$ ,  $d = 1.36$ , 95% CI [0.80, 1.90], SIDC:  $t(22) = 5.18$ ,  $p < .001$ ,  $d = 1.08$ , 95% CI [0.56, 1.59] but a negative effect in the NCATC,  $t(22) = -2.93$ ,  $p = .008$ ,  $d = -0.61$ , 95% CI [-1.05, -0.16],  $F(1,91)$ . Sensitivity analysis (Faul et al., 2007) with  $\alpha = .05$ , Power = .80,  $N = 95$ , and four conditions showed for the detected overall effect for the ANOVA a necessary effect size of  $\eta_p^2 = .11$ .

### Match-mismatch effects

The analysis of the match-mismatch effects revealed the predicted Meaning  $\times$  Group interaction,  $F(3,90)^9 = 3.19$ ,  $p = .027$ ,  $\eta^2 = .10$ , 90% CI [.006, .18] (Figure 2). The sensitivity analysis (Faul et al., 2007) with  $\alpha = .05$ , Power = .80,  $N = 94$ , four conditions, and a correlation between measures of  $r = .31$ , showed for this repeated measures within-between interaction a necessary effect size of  $\eta_p^2 = .04$ . Results showed no main effect of Group (ingroup vs. outgroup),  $F(1,90) = 1.59$ ,  $p = .210$ , or Meaning,  $F(3,90) = 0.82$ ,  $p = .486$ .

When comparing the ingroup match-mismatch effect between the conditions, as predicted, the effect was significantly stronger in the SIDC than in the DIDC,  $t(45) = 2.58$ ,  $p = .012$ ,  $d = 0.75$ , 95% CI [0.16, 1.34]. SIDC and NIDC did not differ significantly,  $t(46) = 1.82$ ,  $p = .072$ ,  $d = 0.53$ , 95% CI [-0.05, 1.10]. All other conditions did not differ, all  $ps > .29$ . The outgroup match-mismatch effect did not differ between any of the conditions, all  $ps > .51$ .

Pairwise comparisons of the match-mismatch effects between ingroup and outgroup within each condition showed that in the SIDC, the ingroup match-mismatch

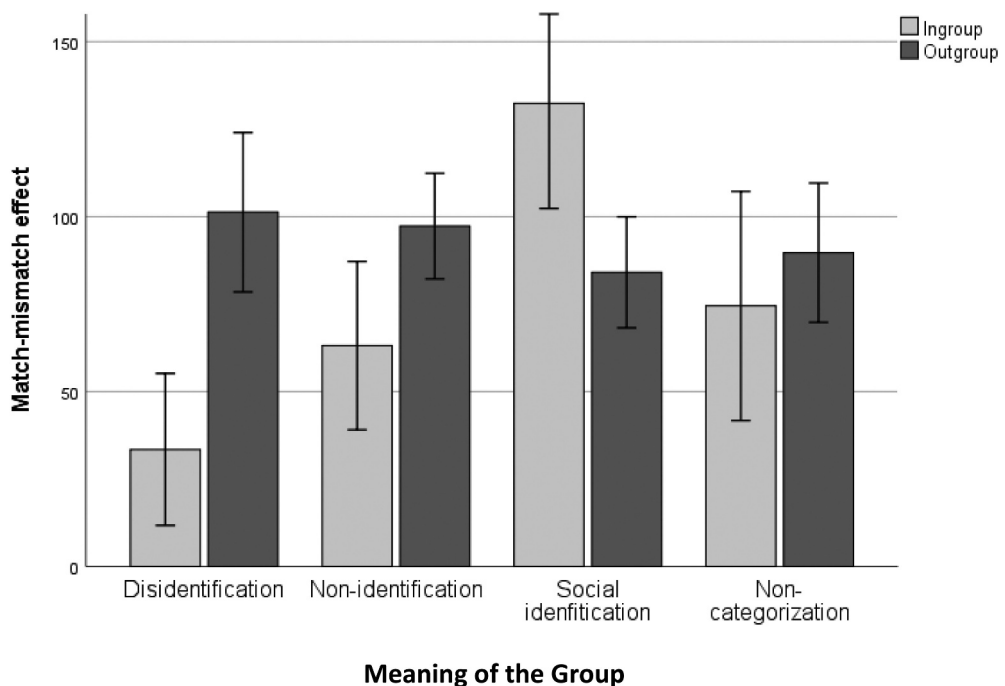


Figure 2. Mean match-mismatch effects (with +/- 1 SE) in Study 2 ( $N = 94$ ).

effect was not significantly stronger than the outgroup effect,  $t(22) = 1.75$ ,  $p = .083$ ,  $d = 0.52$ , 95% CI  $[-0.07, 1.10]$ . In the DIDC, the ingroup match-mismatch effect was significantly weaker than the outgroup effect,  $t(23) = -2.51$ ,  $p = .014$ ,  $d = -0.71$ , 95% CI  $[-0.14, -1.28]$ , indicating a reversed match-mismatch effect. The match-mismatch effects of ingroup and outgroup did not differ significantly between NIDC and NCATC, both  $ps > .19$ .

## Discussion

Study 2 extended the results of Study 1 with the addition of a NCATC. Replicating findings of Study 1, the IAT effect did not differ between the ingroup conditions that held different meanings for the group member (i.e., DIDC, NIDC, and SIDC) but captured differences between ingroup conditions and the condition of no group membership, NCATC. The IAT effect for the conditions with different ingroup meanings was again positive and different from zero, controlling for task order effects. Importantly, if people were not a member of the target groups, the IAT effect was substantially smaller. These results provide initial evidence indicating that identity IATs are more sensitive to self-categorization than to differences in the psychological meaning of an ingroup. As in Study 1, the data support the notion that self-reported social identification and disidentification measures differentiate between non-identification and the respective construct.

For the MMP, as in Study 1, the ingroup match-mismatch effect distinguished the conditions of social identification and disidentification (SIDC and DIDC), whereas the effect in the DIDC was comparable to conditions of non-identification (NIDC) and non-categorization (NCATC). Other than in former research (e.g., Smith et al., 1999; Smith & Henry, 1996), in the SIDC, the ingroup and outgroup match-mismatch effects did not differ significantly. However, results showed a *reversed* effect in the DIDC: the match-mismatch effect was *weaker* for the ingroup compared to the outgroup. Thus, mismatches were facilitated in comparison to matches with the ingroup. This is in line with the definition of disidentification: disidentified members define themselves as *opposite* to their group and are probably used to describing themselves in opposition to ingroup-descriptive traits.

Altogether, Study 2 extended the findings of Study 1 by indicating that the identity IAT mostly captured whether people consider the group to be an ingroup or an outgroup (i.e., self-categorization). Moreover, Study 2 replicated that the MMP differentiates between social identification and disidentification, thus supporting our hypothesis that the MMP captures differences in the psychological meaning of an ingroup. Other than in Study 1, we found no difference between disidentification and non-identification in the ingroup match-mismatch effect, and the ingroup vs. outgroup match-mismatch differences were not found for social identification, but a reversed pattern for disidentification that is worth exploring in another study. Moreover, participants in the NCATC commented on the ambivalent nature of the group of Koreans, as their evaluation and stereotypes strongly differed for North vs. South Korea. Therefore, Study 3 aimed at replicating the results with a different outgroup in the NCATC and an adequate sample size.

### Study 3

Study 3 was a pre-registered online study that tested the hypotheses in a conceptual replication with adequate statistical power.

#### Method

##### Participants

We conducted separate a priori power analyses (Faul et al., 2007) for the predicted effect in the MMP and the identity IAT. A priori power analyses for between-within interaction require the estimation of additional parameters, which comes with added uncertainty and may result in too optimistic sample-size estimates. Following the recommendations of Döring and Bortz (2016, p. 848), we chose a conservative way using a between-subjects ANOVA for the MMP in the power analysis. This way, the planned power will be achieved irrespective of the factual correlation between repeated measurements. This ANOVA (fixed effects, special, main effects and interactions) with  $\alpha = .05$ , Power = .90, and an estimated effect size of  $\eta_p^2 = .09$  (based on the smallest effect size in the pilot data), numerator of the interaction  $df = 3$ , number of groups = 8, for the interaction of meaning (DIDC, NIDC, SIDC, NCATC)  $\times$  group (ingroup vs. outgroup), indicated a necessary sample size of  $N = 148$ . The power analyses for the follow-up pairwise comparisons with a  $t$ -test for independent means (effect size  $d = 0.52$ , based on the smallest effect size in studies 1–2,  $\alpha = .05$ , Power = .90, allocation ratio = 1), and the priori power analyses for the overall effect in the ANOVA to analyze the IAT effect (with  $\alpha = .05$ , Power = .90, and an effect size of  $\eta_p^2 = .32$ ) indicated smaller sample sizes. In case participants are members of the specified outgroups, we added the exclusion criterion of not being Vietnamese or Japanese along with the added exclusion criterion of being a member of the ingroup in all ingroup conditions. To allow for about 20% of exclusions and to power the analyses adequately for the analyses with the weakest effect, we aimed to assess a sample size of  $N = 178$ .

We conducted this study online with Inquisit Web and collected data from 168 participants via Prolific.com. The average study completion time was 25:03 minutes and participants received 5.11£. We excluded  $n = 7$  who did not follow instructions,  $n = 14$  whose data were incomplete in the MMP,  $n = 4$  outliers in the match-mismatch effects, and  $n = 4$  who had more than 10% response times below 300 ms in the IAT. We analyzed data of 137 participants (39 females, 97 males, 1 diverse, 34.4% indicated their age to be in the range of 18–27 years, 40.9% 28–37 years, 16.8% 38–47, 4.4% 48–57, 2.9% 58–67, and 0.7% 68 years or older). Participants were randomly allocated to the conditions: DIDC ( $n = 35$ ), NIDC ( $n = 29$ ), SIDC ( $n = 38$ ), and NCATC ( $n = 35$ ).

##### Procedure and measures

Study 3 followed the procedure of Study 2 with one alteration: In the NCATC, participants referred to the group of Vietnamese instead of Koreans because pretests had demonstrated that this group was similarly neutrally evaluated as the reference outgroup of Japanese. In addition, we included three dichotomous items, one to exclude participants who did not belong to the indicated ingroup (DIDC, NIDC, and SIDC), and two items to

exclude those who belong to the outgroup of Vietnamese (NCATC) and to the reference outgroup of Japanese (see Supplement).

We computed the IAT-D-effect (Spearman Brown coefficient for test halves = .62) and the ingroup and outgroup effects in the MMP and averaged all items of the respective self-reported scales (social identification,  $\alpha = .93$  and disidentification,  $\alpha = .95$ ).

## Results

### Self-reported social identification and disidentification

The MANOVA with Meaning as independent variable and social identification and disidentification as dependent variables demonstrated significant differences between the conditions,  $F(6,266) = 16.77$ ,  $p < .001$ ,  $\eta_p^2 = .27$ , 90% CI [.19, .33] (Table 4). The conditions differed in social identification,  $F(3,137) = 20.52$ ,  $p < .001$ ,  $\eta_p^2 = .32$ , 90% CI [.20, .39]: social identification was higher in the SIDC than in the DIDC,  $t(71) = 4.14$ ,  $p < .001$ ,  $d = 0.97$ , 95% CI [0.48, 1.45], and higher than in the NIDC,  $t(65) = 5.28$ ,  $p < .001$ ,  $d = 1.30$ , 95% CI [0.77, 1.83], and the NCATC,  $t(71) = 7.57$ ,  $p < .001$ ,  $d = 1.77$ , 95% CI [1.23, 2.31]. The DIDC was stronger in social identification than NCATC,  $t(68) = 3.36$ ,  $p = .006$ ,  $d = 0.80$ , 95% CI [0.31, 1.29], but did not differ from NIDC,  $t(62) = 1.32$ ,  $p = 1.000$ ,  $d = 0.33$ , 95% CI [-0.17, 0.83]. NCATC also did not differ from NIDC,  $t(62) = -1.88$ ,  $p = .377$ ,  $d = 0.47$ , 95% CI [-0.97, 0.03]. The conditions also differed in disidentification,  $F(3,137) = 11.51$ ,  $p < .001$ ,  $\eta_p^2 = .21$ , 90% CI [.10, .28]. In the DIDC, participants showed higher disidentification than in the

**Table 4.** Descriptive statistics of the self-reported and indirect measures based on experimental conditions in study 3 ( $N = 137$ ).

	Dis-identification		Social Identification		Non-Identification		Non-categorization	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Social ID</i>	4.16 <sub>ab</sub>	1.39	5.20 <sub>c</sub>	0.85	3.81 <sub>bd</sub>	1.15	3.30 <sub>d</sub>	0.82
<i>DisID</i>	3.73 <sub>a</sub>	1.66	2.05 <sub>b</sub>	0.98	2.82 <sub>bc</sub>	1.17	3.08 <sub>ac</sub>	1.05
<i>MM effect</i>								
Ingroup	0.49 <sub>a</sub>	112.25	88.57 <sub>b</sub>	117.35	62.50 <sub>bc</sub>	90.66	30.15 <sub>ac</sub>	122.23
Outgroup	39.74 <sub>a</sub>	125.16	68.35 <sub>a</sub>	100.57	58.39 <sub>a</sub>	107.30	30.55 <sub>a</sub>	140.69
<i>IAT-D effect</i>	0.39 <sub>a</sub>	0.37	0.47 <sub>a</sub>	0.44	0.36 <sub>a</sub>	0.35	-0.24 <sub>b</sub>	0.29

MM effect = Match-Mismatch Effect Ingroup; ID = Identification; DisID = Dis identification. In the non-categorization condition, ingroup was replaced by the outgroup Vietnamese. Different subscripts within the line indicate statistically significant differences between conditions.

**Table 5.** Descriptive statistics of the match-mismatch effect and the IAT effect in the meta-analysis using a linear-mixed model ( $N = 299$ ).

	Dis-identification		Social Identification		Non-Identification		Non-categorization	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>Match-Mismatch effect</i>								
Ingroup	20.91 <sub>a</sub>	108.02	114.00 <sub>b</sub>	123.49	58.67 <sub>b</sub>	104.84	47.26 <sub>b</sub>	135.61
Outgroup	63.52 <sub>a</sub>	113.11	74.72 <sub>a</sub>	89.68	74.72 <sub>a</sub>	100.22	53.37 <sub>a</sub>	126.98
<i>IAT-D effect</i>	0.50 <sub>a</sub>	0.47	0.58 <sub>a</sub>	0.45	0.53 <sub>a</sub>	0.42	-0.24 <sub>b</sub>	0.33

In the non-categorization condition ingroup was replaced by the out group Koreans/Vietnamese Different subscripts within the line indicate statistically significant differences between conditions.

NIDC,  $t(62) = 2.91$ ,  $p = .025$ ,  $d = 0.73$ , 95% CI [0.22, 1.24] and in the SIDC,  $t(71) = 5.80$ ,  $p < .001$ ,  $d = 1.36$ , 95% CI [0.84, 1.86], but the DIDC did not differ significantly from the NCATC,  $t(68) = 2.19$ ,  $p = .184$ ,  $d = 0.52$ , 95% CI [0.05, 1.00]. The SIDC was significantly lower in disidentification than the NCATC,  $t(71) = -3.57$ ,  $p = .003$ ,  $d = -0.84$ , 95% CI [-1.31, -0.35], but failed to differ significantly from NIDC,  $t(65) = -2.54$ ,  $p = .074$ ,  $d = -0.63$ , 95% CI [-1.12, -0.13]. Taken together, the data confirms the effectiveness of the manipulation and adds to the evidence that only the appropriate measure differentiates indifference from the respective construct.

### Identity IAT effect

In all ingroup conditions, positive IAT effects indicate stronger self-ingroup connections than self-Japanese connections. In the NCATC, positive IAT effects indicate stronger connections of self and Vietnamese than self and Japanese. To test whether the IAT effect is smaller in the NCATC than in the ingroup conditions (i.e., SIDC, DIDC, and NIDC), we conducted an ANOVA with the IAT effect as dependent variable and Meaning as independent variable.

There was a significant difference between conditions,  $F(3,136) = 27.29$ ,  $p < .001$ ,  $\eta_p^2 = .38$ , 90% CI [.26, .46] (Table 4): post-hoc comparisons demonstrated that the IAT effect was lower in the NCATC than in the SIDC,  $t(71) = -8.16$ ,  $d = -1.91$ , 95% CI [-2.46, -1.35], the NIDC,  $t(62) = -6.42$ ,  $d = -1.61$ , 95% CI [-2.18, -1.04], and the DIDC,  $t(68) = -7.13$ ,  $d = -1.70$ , 95% CI [-2.25, -1.15], all  $ps < .001$ . Replicating results from Studies 1 and 2, none of the ingroup conditions differed from each other in the IAT effect, all  $ps = 1.000$ . Sensitivity analysis (Faul et al., 2007) with  $\alpha = .05$ , Power = .80,  $N = 137$ , and four conditions showed for the detected overall effect for the ANOVA a necessary effect size of  $\eta_p^2 = .08$ .

As in Study 1 and 2,  $t$ -tests against 0 showed a positive IAT effect in each of the ingroup conditions, DIDC:  $t(34) = 6.30$ ,  $p < .001$ ,  $d = 1.06$ , 95% CI [0.64, 1.48], NIDC:  $t(28) = 5.59$ ,  $p < .001$ ,  $d = 1.04$ , 95% CI [0.58, 1.49], SIDC:  $t(37) = 6.58$ ,  $p < .001$ ,  $d = 1.07$ , 95% CI [0.66, 1.46], but a negative effect in the NCATC,  $t(34) = -4.74$ ,  $p < .001$ ,  $d = -0.80$ , 95% CI [-1.18, -0.42].

### Match-mismatch effects

Other than expected, the Meaning  $\times$  Group interaction was not significant,  $F(3,133) = 1.06$ ,  $p = .369$ ,  $\eta_p^2 = .02$ , 90% CI [.00, .06] (Figure 3). The sensitivity analysis (Faul et al., 2007) with  $\alpha = .05$ , Power = .80,  $N = 137$ , four conditions, and a correlation between measures of  $r = .23$ , showed a necessary effect size of  $\eta_p^2 = .18$  for this repeated-measures within-between interaction, which was not reached with the data of this study. There was no main effect of Group (ingroup vs. outgroup),  $F(1,133) = .93$ ,  $p = .761$ , but a significant main effect of Meaning,  $F(3,133) = 3.21$ ,  $p = .025$ ,  $\eta^2 = .07$ , 90% CI [.005, .13].

Even though the predicted interaction was not significant, the simple comparisons were in line with the predicted pattern. The ingroup match-mismatch effect in the SIDC was, as predicted, significantly stronger than in the DIDC,  $t(71) = 3.35$ ,  $p = .001$ ,  $d = 0.78$ , 95% CI [0.31, 1.26] and also stronger than in the NCATC,  $t(71) = 2.22$ ,  $p = .028$ ,  $d = .52$ , 95% CI [0.05, 0.99]. In addition, the DIDC had a significantly lower ingroup match-mismatch effect than the NIDC,  $t(62) = -2.20$ ,  $p = .030$ ,  $d = -0.55$ , 95% CI [

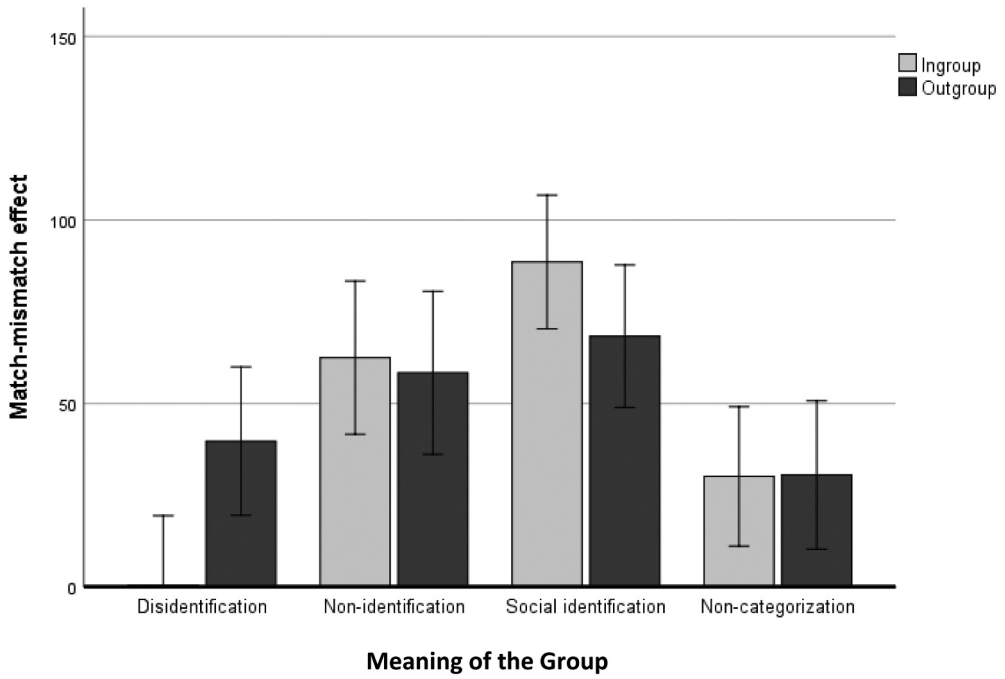


Figure 3. Mean match-mismatch effects (with +/- 1 SE) in Study 3 (N = 137).

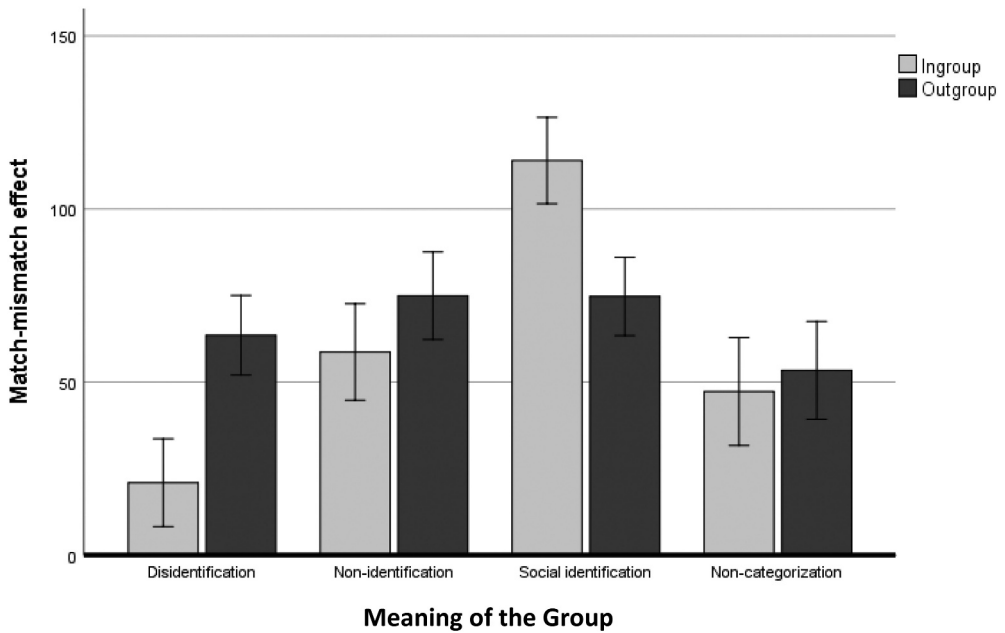


Figure 4. Mean match-mismatch effects (with +/- 1 SE) in the Meta-Analysis (N = 299).

−1.05, −0.05]. All other differences between conditions were not significant, all  $ps > .25$ . The outgroup match-mismatch effects did not differ between any of the conditions, all  $ps \geq .18$ .

The comparison between ingroup and outgroup match-mismatch effect within each condition did not reveal any difference in the SIDC,  $p = .396$ , nor in the other conditions, all  $ps > .11$ .

## Discussion

As predicted, Study 3 replicated that the IAT effect does not differentiate the meanings of the ingroup (i.e., DIDC, NIDC, and SIDC), but indicates whether people self-categorize themselves into a group. As in studies 1 and 2, the social identification and disidentification scales captured differences between the respective construct and non-identification. For the MMP, Study 3 showed the predicted pattern, but the overall interaction between Meaning and the Group (ingroup vs. outgroup) match-mismatch effect was not significant. The direct comparison of the ingroup match-mismatch effect between the SIDC and the DIDC revealed, however, the predicted stronger ingroup effects for social identification, whereas the ingroup match-mismatch effect for disidentification was again comparable to both non-identification and non-categorization. Even though the descriptive pattern is in the predicted direction, the difference between ingroup and outgroup match-mismatch effects within the SIDC was not significant, and there was no reversed effect in the DIDC condition either.

One reason for the lack of evidence for the predicted interaction could be that the effect was too small to be detected with the present sample size, which suggests that the effect is smaller or less robust than anticipated based on Studies 1–2. A crucial difference between Studies 1–2 and Study 3 is the online data collection, which might have added random variance. This notion is supported by the surprisingly high standard deviation of the match-mismatch effects in Study 3. Another difference is that in Study 3, about two thirds of the sample were men, whereas in studies 1–2, two-thirds of the sample were women. In general, women show higher levels of ingroup identification than men (Latrofa et al., 2010; Rudman, 2004; Schmitt et al., 2002). In support of this idea, when comparing Study 3 with Studies 1–2, we found lower levels of ingroup match-mismatch effects, which probably rendered (in combination with a high standard distribution) the overall MMP interaction weaker.

In summary, we found that the IAT differentiates self-categorization from non-categorization but does not capture different meanings of an ingroup. All three studies provide evidence that the self-report measures differentiate between the respective construct and non-identification, whereas the scales that do not explicitly focus on a specific ingroup meaning do not reliably capture differences to non-identification. Regarding the MMP, the evidence is mixed. Whereas the pattern of all data is in the predicted direction, in only two of three studies, the overall interaction between the Meaning condition and ingroup vs. outgroup was significant. All studies found the predicted difference between social identification and disidentification in ingroup match-mismatch effect strength, and two of three studies even found significant differences between the disidentification and non-identification. The ingroup-outgroup difference in the SIDC was, however, only

found once, whereas one study produced a reversed ingroup–outgroup effect in the DIDC. To gain an overall picture of the strength and reliability of the findings, a meta-analysis was conducted across the data sets.

## Meta-Analysis

For the meta-analysis of the IAT effect across all three studies (Table 5), we utilized a linear-mixed effects model (Boedhoe et al., 2019) with Meaning as fixed effects and study as random effects on participants' IAT effects. The variance component analysis suggested a moderate level of heterogeneity between studies (11% of the total variance explained by differences between the studies;  $\sigma_{\text{Study}}^2 = 0.02$ ,  $\sigma_{\text{Residual}}^2 = 0.17$ ).<sup>10</sup>

The analysis showed a significant difference between conditions,  $\chi^2(3) = 124.71$ ,  $p < .001$ . Post-hoc comparisons replicated the smaller IAT effects in the NCATC compared to the other three conditions, all  $ps < .001$ . The IAT effect was lower in the NCATC than in the SIDC, the NIDC, and the DIDC, all  $ps < .001$ . As before, none of the ingroup conditions significantly differed from each other in the IAT effect, all  $ts < 1.28$ . Overall results again showed a positive IAT effect that differed from zero in all ingroup conditions, DIDC:  $t(2.78) = 5.27$ ,  $p = .001$ ,  $d = 0.32$ ; 95% CI [0.43, 2.00], NIDC:  $t(3.05) = 5.63$ ,  $p = .001$ ,  $d = 0.43$ , 95% CI [0.56, 2.09], SIDC:  $t(2.74) = 6.14$ ,  $p = .001$ ,  $d = 0.51$ ; 95% CI [0.61, 2.20], but no IAT effect in the NCATC:  $t(3.60) = -1.46$ ,  $p = .22$ ,  $d = -1.25$ ; 95% CI [-1.08, 0.36].

Additionally, we conducted a meta-analysis for the MMP effects across all three studies, using a linear-mixed effects model with Meaning and Group as fixed effects and study and participants as random effects on participants' responses in the MMP.<sup>11</sup> The variance components analysis suggested a small level of heterogeneity between studies with approximately 2% of the total variance explained by differences between the studies (variance of the study  $\sigma_{\text{Study}}^2 = 300.81$ , variance of the participants  $\sigma_{\text{Participant}}^2 = 2959.82$ , residual variance  $\sigma_{\text{Residual}}^2 = 9421.50$ ).

The analysis showed the predicted interaction between Group Meaning Conditions and Ingroup/Outgroup,  $\chi^2(3) = 16.19$ ,  $p = .001$  (Figure 4). As predicted, the *ingroup* match-mismatch effect was significantly stronger in the SIDC than in the DIDC,  $t(562.95) = 5.57$ ,  $p < .001$ ,  $d = -0.97$ , 95% CI [-1.31, -0.63], and also than in the NIDC,  $t(563.47) = 3.20$ ,  $p = .008$ ,  $d = -0.59$ , 95% CI [-0.95, -0.23], and the NCATC,  $t(554.65) = 3.36$ ,  $p = .005$ ,  $d = 0.67$ , 95% CI [0.28, 1.06]. The *ingroup* match-mismatch effect in the DIDC did not significantly differ from neither the NIDC,  $t(563.40) = -2.06$ ,  $p = .167$ ,<sup>12</sup>  $d = -0.38$ , 95% CI [-0.74, -0.02], nor from the NCATC,  $t(555.05) = -1.48$ ,  $p = .456$ ,  $d = -0.30$ , 95% CI [-0.69, 0.10]. The *outgroup* match-mismatch effect did not differ significantly between conditions, all  $ts < 1.03$ , all  $ps > .73$ .

When comparing ingroup and outgroup match-mismatch effects within each condition, we found in the SIDC that the ingroup (compared to the outgroup) match-mismatch effect is stronger,  $t(299) = 2.70$ ,  $p = .007$ ,  $d = 0.41$ , 95% CI [0.11, 0.70]. In the DIDC, a reversed match-mismatch effect was found, where the outgroup match-mismatch effect was stronger than the ingroup effect  $t(299) = -2.88$ ,  $p = .004$ ,  $d = 0.44$ , 95% CI [-0.74, -0.14]. In the NIDC and the NCATC, ingroup and outgroup match-mismatch effects did not differ significantly, both  $ps > .30$ .

## General discussion

Three studies along with their meta-analysis tested whether the most established indirect measures for social identity, the identity IAT and the MMP, distinguish social identification from disidentification, two qualitatively different meanings of an ingroup.

As expected, results consistently showed that the IAT differentiated self-categorization from non-categorization but did not capture different meanings of the ingroup: social identification, disidentification, and non-identification conditions all resulted in IAT effects of comparable size. Only participants who did not belong to any of the groups showed a substantially smaller IAT effect than all ingroup conditions. The MMP, as expected, differentiated social identification from disidentification: across all data sets, the ingroup match–mismatch effect was stronger for group members who identified compared to those who disidentified with their ingroup. Additionally, the meta-analysis confirmed the difference in match-mismatch effects between ingroups and outgroups (that constitute the baseline of the effect) for identified group members. Moreover, it showed initial evidence for a reversed ingroup-outgroup match-mismatch effect for the disidentified group members. Finally, the data contribute to the finding that the self-reported social identification and disidentification scales measure the respective construct reliably, as they specifically differentiate between social identification/disidentification and non-identification.

### *The interpretation of the identity IAT and MMP*

Identity IATs and the MMP have often been interpreted as measures of social identification (e.g., Aidman & Carroll, 2003; Chen et al., 2021; Coats et al., 2000; Cvencek et al., 2011; Devos & Torres, 2007; Hawkins & Nosek, 2012; Kirby et al., 2021; Rudman & Goodwin, 2004; Smith & Henry, 1996). Our findings demonstrate, however, that this interpretation might lead to false conclusions, especially if participants vary in the meaning that they attach to their ingroup. More precisely, our findings reveal that if the identity IAT is taken as a measure of social identification, both disidentified and non-identified group members would figure as identified group members, whereas only participants who do not self-categorize into the group would be taken as non-identified. This interpretation might be unproblematic when the identity IAT is primarily used in contexts where ingroups have a positive meaning, but will lead to false conclusions when used for ingroups containing a neutral or negative meaning. For example, applying identity IATs for stigmatized ingroups (e.g., Grover et al., 2003; Kirby et al., 2021), drug or alcohol-related identities (Albery et al., 2024; Chen et al., 2021; Gray et al., 2011) may provide positive identity IAT effects that, instead of identification, are actually produced by disidentification. As social identification and disidentification have different antecedents and consequences, research might reveal odd or no relations to other constructs if disidentified group members are taken for identified group members based on an identity IAT.

In contrast, our findings suggest that the MMP can be used to differentiate social identification from disidentification: in the MMP, disidentified group members show lower match-mismatch effects with the ingroup than identified group members. Thus, from the MMP measure, researchers can induce the quality of the relation that ties an individual to an ingroup. When the MMP is, however, interpreted as a measure of social identities (as

suggested by Smith & Henry, 1996; Smith et al., 1997) or self-categorization, and researchers are unaware of the nature of the relation between self and ingroup, the measure might lead to the (false) conclusion that disidentified participants have not formed meaningful social identities at all. Taking both disidentified and non-identified group members as individuals without a social identity for the specific group would mean to trivialize a potentially dangerous negative ingroup relation, as undetected disidentified group members can easily harm the ingroup from within (e.g., Becker & Tausch, 2014; De Vreeze & Matschke, 2017; 2019). Thus, our data underline that considering the exact nature of the meaning of the group is necessary when using the MMP, as it captures more than the mere, unqualified connection between self and ingroup, but a self-description in line/in contrast with the group's characteristics.

It is an interesting question whether the MMP and the identity IAT capture different underlying cognitive processes, and thus whether one measure is more prone to be influenced by association while the other measure is more influenced by relational inferences or propositional reasoning (de Houwer et al., 2020; Gawronski & Bodenhausen, 2006). More specifically, the observation that the identity IAT primarily captures group membership and is insensitive to the quality of that membership could be interpreted as evidence for the identity IAT to capture unqualified associations between the self and the group. Upon closer reflection, however, both, an associative and a propositional theoretical perspective seem compatible with the empirical observations. From an association-based perspective, one may assume that associations between self and group result from repeated co-activation (e.g., Hutchison, 2003; Lucas, 2000, for reviews), which may occur for identified and disidentified members alike. From a proposition-based perspective, one may assume that inferences and reasoning impact results on indirect measures (Charlesworth et al., 2020; de Houwer, 2009; Kurdi & Banaji, 2019; Mitchell et al., 2009). Given that the concept of disidentification presupposes that the disidentified person still self-categorizes as a group member, it follows that this knowledge will affect identity IAT scores. Because the procedural nature of the identity IAT explicitly and only requires participants to engage in factual categorizations (i.e., evaluating the truth of "X belongs to category Y" propositions), evaluations of the quality of group membership can be expected to play a minor role for responding. In essence, both accounts support our hypothesis that identity IATs will not differentiate between ingroup identification and disidentification.

A more differentiated picture emerges for the results obtained with the MMP. Here, the match-mismatch effect for the ingroup proved substantially reduced for disidentified vs. identified participants. This finding results both from faster responses in affirming traits that do not belong to the ingroup and negating traits that do belong to the ingroup (i.e., mismatches) as well as being slower in affirming traits that do belong to the ingroup and negating traits that do not belong to the ingroup (i.e., matches) than identified participants. Clearly, for affirmative responses, results can be easily understood from both association and proposition perspectives alike. For example, disidentified participants can be assumed to have formed stronger associations between the self and non-ingroup traits than identified participants. Likewise, they can be assumed to have formed a memory representation of the proposition that these traits apply to them. For the negated responses, however, a propositional perspective seems to explain the observations more readily than an associative perspective. The case of negating that an ingroup

trait applies to the self may serve as an example (e.g., a disidentified German negates being punctual). Why would being disidentified increase the speed of negating the trait? From a propositional perspective, one could assume that repeatedly negating to possess ingroup traits results in the formation of a memory representation of this negation that can be quickly retrieved (e.g., de Houwer, 2014; Zou et al., 2018). From an associative perspective, it is more ambiguous which associative memory structure could support fast negations of ingroup traits. One possibility would be mutually activating links between the self, the trait, and a “negation tag” (cf. Petty et al., 2007). In the MMP, seeing the trait might activate all components readily, thereby facilitating the judgment. At the same time, given that the mere activation of the three components includes high accessibility of traits that do not apply to the self, this would create an ambiguous state (Petty et al., 2007) that needs to be resolved by additional processes, thereby potentially slowing down responses. Irrespective of whether identity IATs or the MMP are assessing more or less associative versus propositional processes, the present data demonstrate that the identity IAT assesses self-categorization whereas the MMP captures the quality of this self-categorization in terms of disidentifying or identifying with the groups.

Taken together, our findings contribute to a better understanding of the aspect of social identities that the indirect measures assess. The findings are the first to demonstrate that meanings of ingroups are reflected in the MMP but not in identity IATs. Thereby, future research can make use of identity IATs and MMP more purposefully.

### **Contribution to the identity IAT**

Identity IATs, as implemented in the present research, are widely used to assess social identity (e.g., Aidman & Carroll, 2003; Devos et al., 2010; Theodoridis, 2013, 2017). However, other researchers have suggested that those IATs assess self-categorization instead of social identification (Cadinu & Galdi, 2012). The present research is the first to systematically test this suggestion. It clearly demonstrated that identity IATs the way they are currently used are rather measuring the strength of self-categorization than social identification. That said, self-categorization itself has relevant outcomes such as ingroup bias (Billig & Tajfel, 1973). Thus, it is not less relevant what the identity IAT is tapping into. Instead, knowing what the typical identity IATs are assessing contributes to specifying theories and helps to understand social phenomena.

The evidence that the currently implemented identity IATs are mostly capturing self-categorization can also inspire future research on IAT developments. By adding more content to the exemplars used in identity IATs, by changing the IAT labels to be more comprehensive of the groups, or by combining a self-categorization IAT with a group attitude IAT (Glaser & Knowles, 2008), the IAT may become a useful tool to also assess variations in the meaning of ingroups to the self.

Notably, the IAT effects across conditions in the present studies tended to correlate positively with self-reported scales of social identification and negatively with disidentification. This finding suggests that some variance in the self-reported measures captures how strongly people categorize themselves into a social group.

Furthermore, the current IATs did not show null effects when participants were not members of any of the two social groups. This indicates that participants associated themselves stronger with one outgroup than with another one, which was reflected in

the present IAT effects in the non-categorization conditions. What specifically is driving those effects could be investigated in future research.

### *Contribution to the MMP*

Traditionally, the MMP has been used to measure the inclusion of persons or ingroups into the self-concept (Aron et al., 1991; Smith & Henry, 1996; Smith et al., 1997). The present data is the first to demonstrate that the measure differentiates different meanings of ingroups. As the method allows participants to describe themselves and the ingroup, they can describe themselves as similar (i.e., social identification) or in contrast (i.e., disidentification) to the ingroup. As this perceived overlap versus opposition between self and group is the central cognitive distinction between social identification and disidentification, the MMP suits to differentiate these different types of ingroup meanings.

The reversed pattern of the ingroup-outgroup difference for disidentified group members is another novelty of our data. It demonstrates that disidentification channels the rejection of ingroup-prototypical characteristics and inhibits their acceptance. Future research should explore whether this reversed pattern also occurs when the MMP is used to measure interpersonal relations (e.g., Aron et al., 1991). So far, the measure has been applied to close relationships that were positive in nature. Based on our findings, it is likely that the MMP produces a reversed pattern for relevant close others to whom one has a negative relationship.

The traditional comparison between ingroup and outgroup match-mismatch effects was, however, partly challenged by our data. Even though the meta-analysis confirmed for identified group members that the ingroup effect is stronger than the outgroup effect, the pattern was not entirely robust in all studies. One reason for this inconsistency could be that most former research used samples of psychology students (e.g., Smith et al., 1999; Smith & Henry, 1996) or other samples with many women (e.g., Tropp & Wright, 2001). As women show stronger levels of ingroup identification (Latrofa et al., 2010; Rudman, 2004; Schmitt et al., 2002), these samples might have overestimated the effect. Another difference between former research and our studies is the ingroup choice that we allowed participants to make, while in former studies, participants were given a specific ingroup (e.g., a fraternity, gender, or race). Choosing the group should, on the one hand, strengthen the bond to the group, but on the other hand, it might also increase the variance of social identification with the groups in our data. Altogether, future research should test the boundary conditions of the MMP more systematically.

### *Contribution to social identity research*

The current data contribute to the debate on disidentification as an asset to social psychological research. On the one hand, critics of the construct argue that the critical variance that has the most predictive value for ingroup-related phenomena is the difference between high and low levels of social identification. Research has, indeed, not always been able to differentiate social identification from disidentification (e.g., Gligorić & Obradović, 2024). Criticism of the construct receives support by our finding that neither the IAT, nor the MMP reliably differentiates between non-identification and disidentification. Even though two of three studies found this difference in the MMP, the

meta-analysis did not reveal overall support for the distinction. Critics may argue that if that is the case, the differentiation of non-identification and disidentification is irrelevant.

On the other hand, advocates of the construct argue that adding the dimension between indifference (non-identification) and disidentification allows a more complete picture of potential ingroup relations. Research has indeed found that using both constructs reveals relations between the self-ingroup relation and other constructs, that might otherwise be undetected, because the measures are especially sensitive to negative and positive group-related income and outcome means, respectively (e.g., Becker & Tausch, 2014; De Vreeze & Matschke, 2017; 2019; Elsbach & Bhattacharya, 2001; Ikegami & Ischida, 2007; Kreiner & Ashforth, 2004; Matschke & Sassenberg, 2010a, 2010b; Matschke et al., 2023; Zou et al., 2008). In line with the first argument, the current data contributes to the body of evidence demonstrating that differences between non-identification and disidentification are not necessarily captured on self-reported social identification measures, but need disidentification measures to be detected (e.g., Becker & Tausch, 2014). In line with the second argument, the reversed match-mismatch effect, where the ingroup-prototypical characteristics are more strongly rejected than the outgroup-prototypical characteristics, was (in the meta-analysis) only found for disidentified, but not for non-identified members. Thus, the self-description was affected more strongly by a negative self-ingroup relation than by an irrelevant relation. In fact, the finding that the self-ingroup overlap drops below the level of the self-outgroup overlap is genuinely in line with the definition of disidentification and demonstrates that disidentification contributes to the understanding of self-ingroup related phenomena.

## Conclusion

Indirect measures hold their established place among measures of social identities. The present research demonstrates that while using identity IATs and the MMP, researchers should be aware of the different aspects of social identities that they assess and keep differences in specific self-ingroup relations in mind that produce more or less specific patterns in these measures. One main message from the present research is that researchers need to watch out which fish they want to catch and use the right fishing rod. Additionally, the knowledge of what measures are assessing is crucial for theory development and its refinement on the self-ingroup relationships.

## Notes

1. In their expanded model of organizational identification, Kreiner and Ashforth (2004) suggest 4 types of relations to an ingroup: identification, disidentification, neutral identification, and ambivalent identification. As the current research is the first to test whether indirect measures assess these differences at all, we focus on social identification and disidentification, and add, as a third condition, non-identification for explorative purposes that is neutral identification in terms of the Kreiner and Ashforth model.
2. We conducted another study online that we had set it up in Javascript unknowingly, at that time, that this software was not suited for assessing reaction times. Resulting random error times (due to potential browsers that people may have used to complete the study) lay within the same order of magnitude than the expected reaction time differences based on

conditions. Therefore, we considered this study to be invalid to test the sensitivity of reaction time paradigms and did not report it.

3. The items validate the findings of self-report measures by differentiating the conditions and serve as an extra-manipulation check (see Supplement).
4. see Supplement for all exploratory variables included in Studies 1 and 2.
5. In Studies 1 and 2, we presented participants with 16 items on social identification. One item ("I think that this group [ingroup] has a lot to be proud of") was excluded from the scale, because scale validation shows that this item is not suited for German samples (Roth & Mazziotta, 2015).
6. Note that former research has found weak match-mismatch effects for irrelevant outgroups, too (Otten & Epstude, 2006; Sassenberg & Matschke, 2010). Irrelevant outgroups also produce a difference between matches and mismatches in the reaction times of self-descriptions, but only for traits that were not self-descriptive (Tropp & Wright, 2001). This evidence suggests that the match-mismatch measure is affected by other aspects than ingroup identification. The measure seems partly driven by the same general valence of the self (because of the striving for positive self-esteem) and an irrelevant outgroup (which might be described positively for political correctness). In line with this argument, when differentiated into valent and neutral traits, the match-mismatch effect for the outgroup disappears for neutral traits, but not for valent traits (Sassenberg & Matschke, 2010). For the current research questions, it is important to keep in mind that the baseline of the effect is not zero. Instead, a match-mismatch effect comparison between ingroup and outgroup can serve as a baseline.
7. For all studies, additional Bayes testing of the null hypothesis showed the same results pattern (see Supplement).
8. see Supplement for additional exploratory variables included in Study 2.
9. Two participants had missing data in the MMP. Therefore, analyses are based on  $n = 94$ .
10. This heterogeneity between studies is mostly due to Study 1 not counterbalancing the test blocks while Studies 2 and 3 counterbalance the blocks resulting in a stronger IAT effect in Study 1 compared to Studies 2 and 3 driven by learning effects.
11. When conducting a repeated measures analysis without random effects, the results pattern is comparable to the reported analysis with one difference: The DIDC differs significantly from the NIDC, see Supplement.
12. Even though the confidence interval does not include zero, due to Tukey corrections in multiple comparisons, the  $p$ -value is not significant.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the European Social Fund grand number: [01PF 08005 C], the Deutsche Forschungsgemeinschaft Grand number: [MA 7236 1/1], and by seed funding of the Institute of Psychology of the University of Wuerzburg and the Technical University of Dresden.

## Data availability statement

Anonymized data of Study 3 are available on OSF (for now the reviewer link: <https://osf.io/wxmkc/files/osfstorage>). The data of Studies 1 and 2 are not accessible in a public repository because participants did not give their explicit consent to make the data public. The data are available for research purposes from the corresponding author upon request.

## References

- Aidman, E. V., & Carroll, S. M. (2003). Implicit individual differences: Relationships between implicit self-esteem, gender identity, and gender attitudes. *European Journal of Personality*, *17*(1), 19–36. <https://doi.org/10.1002/per.465>
- Albery, I. P., Long, N., Kelly, I., Frings, D., & Morris, J. (2024). Does identity as a drinker predict problem recognition motivation in harmful drinkers? *Social Science & Medicine*, *351*, 116967. <https://doi.org/10.1016/j.socscimed.2024.116967>
- Aron, A., Aron, E. N., Tudor, M., & Nelson, G. (1991). Close relationships as including other in the self. *Journal of Personality and Social Psychology*, *60*(2), 241–253. <https://doi.org/10.1037/0022-3514.60.2.241>
- Ashforth, B. E., Joshi, M., Anand, V., & O'Leary-Kelly, A. M. (2013). Extending the expanded model of organizational identification to occupations. *Journal of Applied Social Psychology*, *43*(12), 2426–2448. <https://doi.org/10.1111/jasp.12190>
- Barreto, M., & Ellemers, N. (2003). The effects of being categorized: The interplay between internal and external social identities. *European Review of Social Psychology*, *14*(1), 139–170. <https://doi.org/10.1080/10463280340000045>
- Becker, J. C., & Tausch, N. (2014). When group memberships are negative: The concept, measurement, and behavioral implications of psychological disidentification. *Self and Identity*, *13*(3), 294–321. <https://doi.org/10.1080/15298868.2013.819991>
- Becker, J. C., Tausch, N., Spears, R., & Christ, O. (2011). Committed dis(s)idents: Participation in radical collective action fosters disidentification with the broader in-group but enhances political identification. *Personality and Social Psychology Bulletin*, *37*(8), 1104–1116. <https://doi.org/10.1177/0146167211407076>
- Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, *36*(1), 27–52. <https://doi.org/10.1002/ejsp.2420030103>
- Boedhoe, P. S. W., Heymans, M. W., Schmaal, L., Abe, Y., Alonso, P., Ameis, S. H., Anticevic, A., Arnold, P. D., Batistuzzo, M. C., Benedetti, F., Beucke, J. C., Bollettini, I., Bose, A., Brem, S., Calvo, A., Calvo, R., Cheng, Y., Cho, K. I. K., Ciullo, V. . . . Twisk, J. W. R. (2019). An empirical comparison of meta- and mega-analysis with data from the ENIGMA obsessive-compulsive disorder working group. *Frontiers in Neuroinformatics*, *12*, 102. <https://doi.org/10.3389/fninf.2018.00102>
- Brewer, M. B., & Pickett, C. (1999). Distinctiveness motives as a source of the social self. In T. R. Tyler, R. M. Kramer, & O. P. John (Eds.), *The psychology of the social self* (pp. 71–87). Lawrence Erlbaum Associates Publishers.
- Cadinu, M., & Galdi, S. (2012). Gender differences in implicit gender self-categorization lead to stronger gender self-stereotyping by women than by men. *European Journal of Social Psychology*, *42*(5), 546–551. <https://doi.org/10.1002/ejsp.1881>
- Charlesworth, T. E. S., Kurdi, B., & Banaji, M. R. (2020). Children's implicit attitude acquisition: Evaluative statements succeed, repeated pairings fail. *Developmental Science*, *23*, e12911. <https://doi.org/10.1111/desc.12911> (3)
- Chen, J., Zhang, M., Zhou, J., Li, X., Zhang, F., & Shen, M. (2021). Implicit and explicit self-identification as a drug user in people who used heroin and methamphetamine [original research]. *Frontiers in Psychology*, *12*. <https://doi.org/10.3389/fpsyg.2021.685110>
- Clifton, R. L., Rowe, A. T., Banks, D. E., Ashburn-Nardo, L., & Zapolski, T. C. B. (2021). Examining the effects of implicit and explicit racial identity on psychological distress and substance use among black young adults. *Experimental and Clinical Psychopharmacology*, *29*(5), 479–486. <https://doi.org/10.1037/pha0000489>
- Coats, S., Smith, E. R., Claypool, H. M., & Banner, M. J. (2000). Overlapping mental representations of self and in-group: Reaction time evidence and its relationship with explicit measures of group identification. *Journal of Experimental Social Psychology*, *36*(3), 304–315. <https://doi.org/10.1006/jesp.1999.1416>
- Cummins, J., Lindgren, K. P., & Houwer, J. D. (2021). On the role of (implicit) drinking self-identity in alcohol use and problematic drinking: A comparison of five measures. *Psychology of Addictive*

- Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, 35(4), 458. <https://doi.org/10.1037/adb0000643>
- Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2012). Balanced identity theory review of evidence for implicit consistency in social cognition. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 157–177). Guilford Press.
- Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2016). Implicit measures for preschool children confirm self-esteem's role in maintaining a balanced identity. *Journal of Experimental Social Psychology*, 62, 50–57. <https://doi.org/10.1016/j.jesp.2015.09.015>
- Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-gender stereotypes in elementary school children [article]. *Child Development*, 82(3), 766–779. <https://doi.org/10.1111/j.1467-8624.2010.01529.x>
- Cvencek, D., Meltzoff, A. N., & Kapur, M. (2014). Cognitive consistency and math-gender stereotypes in Singaporean children. *Journal of Experimental Child Psychology*, 117, 73–91. <https://doi.org/10.1016/j.jecp.2013.07.018>
- Cvencek, D., Meltzoff, A. N., Maddox, C. D., Nosek, B. A., Rudman, L. A., Devos, T., Dunham, Y., Baron, A. S., Steffens, M. C., Lane, K., Horcajo, J., Ashburn-Nardo, L., Quinby, A., Srivastava, S. B., Schmidt, K., Aidman, E., Tang, E., Farnham, S., Mellott, D. S., & Greenwald, A. G. (2021). Meta-analytic Use of balanced identity theory to validate the implicit association test. *Personality & Social Psychology Bulletin*, 47(2), 185–200. <https://doi.org/10.1177/0146167220916631>
- Deaux, K. (1996). Social identification. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 777–798). Guilford Press.
- de Houwer, J. D. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, 37(1), 1–20. <https://doi.org/10.3758/lb.37.1.1>
- de Houwer, J. D. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, 8(7), 342–353. <https://doi.org/10.1111/spc3.12111>
- de Houwer, J. D., Van Dessel, P., & Moran, T. (2020). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. *Advances in experimental social psychology*, 61, 127–183. <https://doi.org/10.1016/bs.aesp.2019.09.004>
- Devos, T. (2006). Implicit bicultural identity among Mexican American and Asian American college students. *Cultural Diversity and Ethnic Minority Psychology*, 12(3), 381–402. <https://doi.org/10.1037/1099-9809.12.3.381>
- Devos, T., Gavin, K., & Quintana, F. J. (2010). Say “adios” to the American dream? The interplay between ethnic and national identity among latino and Caucasian Americans. *Cultural Diversity and Ethnic Minority Psychology*, 16(1), 37–49. <https://doi.org/10.1037/a0015868>
- Devos, T., & Torres, J. A. C. (2007). Implicit identification with academic achievement among Latino college students: The role of ethnic identity and significant others. *Basic and Applied Social Psychology*, 29(3), 293–310. <https://doi.org/10.1080/01973530701503432>
- De Vreeze, J., & Matschke, C. (2017). Keeping up appearances: Strategic information exchange by disidentified group members. *PLOS ONE*, 12(4), e0175155. <https://doi.org/10.1371/journal.pone.0175155>
- de Vreeze, J., & Matschke, C. (2019). Don't put me in this group assignment to non-preferred groups increases disidentification and a preference for negative ingroup information. *Social Psychology*, 50(2), 80–93. <https://doi.org/10.1027/1864-9335/a000363>
- de Vreeze, J., Matschke, C., & Cress, U. (2018). Neither fish nor fowl: A perceived mismatch in norms and values between oneself, other students, and people back home undermines adaptation to university. *British Journal of Social Psychology*, 57(3), 684–702. <https://doi.org/10.1111/bjso.12253>
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation*. Springer.
- Egloff, B., & Schmukle, S. (2002). Predictive validity of an implicit association test for assessing anxiety. *Journal of Personality and Social Psychology*, 83(6), 1441–1455. <https://doi.org/10.1037/0022-3514.83.6.1441>
- Elsbach, K. D., & Bhattacharya, C. B. (2001). Defining who you are by what you're not: Organizational disidentification and the national rifle association. *Organizational Science*, 12, 393–413. <https://doi.org/10.1287/orsc.12.4.393.10638>

- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\* power, 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/bf03193146>
- Gawronski, B. (2019). Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Science*, 14(4), 574–595. <https://doi.org/10.1177/1745691619826015>
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, 132(5), 692.
- Gawronski, B., Houwer, J. D., & Sherman, J. W. (2020). Twenty-five years of research using implicit measures. *Social Cognition*, 38(Supplement), 1–25. <https://doi.org/10.1521/soco.2020.38.supp.s1>
- Glaser, J., & Knowles, E. D. (2008). Implicit motivation to control prejudice. *Journal of Experimental Social Psychology*, 44(1), 164–172. <https://doi.org/10.1016/j.jesp.2007.01.002>
- Gligorić, V., & Obradović, S. (2024). Active rejection or passive indifference? Mixed-methods evidence on national (dis) identification. *Group Processes & Intergroup Relations*, 27(7), 1519–1540. <https://doi.org/10.1177/13684302241229981>
- Gray, H. M., LaPlante, D. A., Bannon, B. L., Ambady, N., & Shaffer, H. J. (2011). Development and validation of the alcohol identity implicit associations test (AI-IAT). *Addictive Behaviors*, 36(9), 919–926. <https://doi.org/10.1016/j.addbeh.2011.05.003>
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, 109(1), 3–25. <https://doi.org/10.1037/0033-295X.109.1.3>
- Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, 79(6), 1022–1038. <https://doi.org/10.1037/0022-3514.79.6.1022>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197–216. <https://doi.org/10.1037/0022-3514.85.2.197>
- Grover, V. P., Keel, P. K., & Mitchell, J. P. (2003). Gender differences in implicit weight identity. *International Journal of Eating Disorders*, 34(1), 125–135. <https://doi.org/10.1002/eat.10167>
- Hamstra, M. R. W., Sassenberg, K., van Yperen, N. W., Wisse, B., & Rietzschel, E. F. (2015). Regulatory fit buffers against disidentification from groups. *Motivation Science*, 1(3), 184–201. <https://doi.org/10.1037/mot0000019>
- Hawkins, C. B., & Nosek, B. A. (2012). Motivated independence? Implicit party identity predicts political judgments among self-proclaimed independents. *Personality and Social Psychology Bulletin*, 38(11), 1437–1452. <https://doi.org/10.1177/0146167212452313>
- Hughes, E. A., & Smith, J. R. (2025). Measuring associations among British national identification, group norms and social distancing behaviour during the COVID-19 pandemic: Testing a social Identity Model of Behavioural associations (SIMBA). *British Journal of Social Psychology*, 64(2), e12862.
- Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, 10(4), 785–813. <https://doi.org/10.3758/bf03196544>
- IBM Corp. (2023). *IBM SPSS statistics for windows* (version 29.0). Computer software.
- Ikegami, T. (2010). Precursors and consequences of in-group disidentification: Status system beliefs and social identity. *Identity an International Journal of Theory and Research*, 10(4), 233–253. <https://doi.org/10.1080/15283488.2010.523590>
- Ikegami, T., & Ischida, Y. (2007). Status hierarchy and the role of disidentification in discriminatory perception of outgroups<sup>1</sup>. *Japanese Psychological Research*, 49(2), 136–147. <https://doi.org/10.1111/j.1468-5884.2007.00340.x>
- Kelley, K. (2023). *MBESS: Methods for the behavioral, educational, and social sciences* (version 4.9.3) [R package]. <https://CRAN.R-project.org/package=MBESS12>

- Kirby, T. A., Merritt, S. K., Baillie, S., Malahy, L. W., & Kaiser, C. R. (2021). Combating bisexual erasure: The correspondence of implicit and explicit sexual identity. *Social Psychological and Personality Science*, 12(8), 1415–1424. <https://doi.org/10.1177/1948550620980916>
- Kreiner, G. E., & Ashforth, B. E. (2004). Evidence toward an expanded model of organizational identification. *Journal of Organizational Behavior*, 25(1), 1–27. <https://doi.org/10.1002/job.234>
- Kurdi, B., & Banaji, M. R. (2019). Attitude change via repeated evaluative pairings versus evaluative statements: Shared and unique features. *Journal of Personality and Social Psychology*, 116(5), 681–703. <https://doi.org/10.1037/pspa0000151>
- Lakens, D. (2014, June 07). *Calculating confidence intervals for Cohen's d and eta-squared using SPSS, R, and Stata. The 20% Statistician*. <https://daniellakens.blogspot.com/2014/06/calculating-confidence-intervals-for.html>
- Latrofa, M., Vaes, J., Cadinu, M., & Carnaghi, A. (2010). The cognitive representation of self-stereotyping. *Personality and Social Psychology Bulletin*, 36(7), 911–922. <https://doi.org/10.1177/0146167210373907>
- Leach, C. W., van Zomeren, M., Zebel, S., Vliek, M. L. W., Pennekamp, S. F., Doosje, B., Ouwerkerk, J. W., & Spears, R. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology*, 95(1), 144–165. <https://doi.org/10.1037/0022-3514.95.1.144>
- Loughnane, J., Roth, J., Rauner, M., & Strack, F. (2021). From Student to employee: Group compatibility predicts group identification and intergroup attitudes. *Identity*, 21(2), 131–143. <https://doi.org/10.1080/15283488.2021.1896522>
- Loughnane, J., Roth, J., & van Tilburg, W. (2023). A cognitive balance approach to understanding intergroup attitudes in post-Brexit Northern Ireland. *British Journal of Social Psychology*, 62(4), 1798–1816. <https://doi.org/10.1111/bjso.12656>
- Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, 7(4), 618–630. <https://doi.org/10.3758/bf03212999>
- Machery, E. (2022). Anomalies in implicit attitudes research. *Wiley Interdisciplinary Reviews Cognitive Science*, 13(1), e1569. <https://doi.org/10.1002/wcs.1569>
- Matschke, C., de Vreeze, J., & Cress, U. (2023). Social identities and the achievement gap: Incompatibility between social class background and student identity increases student disidentification, which decreases performance and leads to higher dropout rates. *British Journal of Social Psychology*, 62(1), 161–180. <https://doi.org/10.1111/bjso.12563>
- Matschke, C., & Sassenberg, K. (2010a). The supporting and impeding effects of group-related approach and avoidance strategies on newcomers' psychological adaptation. *International Journal of Intercultural Relations*, 34(5), 465–474. <https://doi.org/10.1016/j.ijintrel.2010.04.008>
- Matschke, C., & Sassenberg, K. (2010b). Does rejection lead to disidentification? The role of internal motivation and avoidance strategies. *European Journal of Social Psychology*, 40(6), 891–900. <https://doi.org/10.1002/ejsp.756>
- Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am not guilty" vs "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, 40(4), 433–449. <https://doi.org/10.1016/j.jesp.2003.07.008>
- Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, 32(2), 183–198. <https://doi.org/10.1017/s0140525x09000855>
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality & Social Psychology*, 83(1), 44–59. <https://doi.org/10.1037/0022-3514.83.1.44>
- Otten, S., & Epstude, K. (2006). Overlapping mental representations of self, ingroup, and outgroup: Unraveling self-stereotyping and self-anchoring. *Personality and Social Psychology Bulletin*, 32(7), 957–969. <https://doi.org/10.1177/0146167206287254>
- Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The meta-cognitive model (MCM) of attitudes: Implications for attitude measurement, change, and strength. *Social Cognition*, 25(5), 657–686. <https://doi.org/10.1521/soco.2007.25.5.657>

- Postmes, T., Haslam, S. A., & Jans, L. (2012). A single-item measure of social identification: Reliability, validity, and utility. *British Journal of Social Psychology, 52*(4), 597–617. <https://doi.org/10.1111/bjso.12006>
- R Core Team. (2021). *R: A language and environment for statistical computing* (version 4.0.5) [computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Roth, J., & Mazziotta, A. (2015). Adaptation and validation of a German multidimensional and multicomponent measure of social identification. *Social Psychology, 46*(5), 277–290. <https://doi.org/10.1027/1864-9335/a000243>
- Roth, J., Steffens, M. C., & Vignoles, V. L. (2018). Group membership, group change, and intergroup attitudes: A recategorization Model based on cognitive consistency principles [hypothesis and theory]. *Frontiers in Psychology, 9*(479). <https://doi.org/10.3389/fpsyg.2018.00479>
- Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science, 13*(2), 79–82. <https://doi.org/10.1111/j.0963-7214.2004.00279.x>
- Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic In-group Bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology, 87*(4), 494–509. <https://doi.org/10.1037/0022-3514.87.4.494>
- Sassenberg, K., & Matschke, C. (2010). The impact of exchange programs on the integration of the host-group into the self-concept. *European Journal of Social Psychology, 40*(1), 148–159. <https://doi.org/10.1002/ejsp.621>
- Schimmack, U. (2021). The implicit association test: A method in search of a construct. *Perspectives on Psychological Science, 16*(2), 396–414. <https://doi.org/10.1177/1745691619863798>
- Schmitt, M. T., Branscombe, N. R., Kobrynowicz, D., & Owen, S. (2002). Perceiving discrimination against one's gender group has different implications for well-being in women and men. *Personality and Social Psychology Bulletin, 28*(2), 197–210. <https://doi.org/10.1177/0146167202282006>
- Smith, E. R. (2002). Overlapping mental representations of self and group evidence and implications. In J. P. Forgas & U. D. Williams (Eds.), *The social self. Cognitive, interpersonal and intergroup perspectives* (pp. 21–35). Psychology Press.
- Smith, E. R., Coats, S., & Walling, D. (1997). *Correlations of RT facilitation effects with scale measuring group identification [data set]*. Department of Psychological Sciences, Purdue University. <https://doi.org/10.1037/e413782005-656>
- Smith, E. R., Coats, S., & Walling, D. (1999). Overlapping mental representations of self, in-group, and partner: Further response time evidence and a connectionist model. *Personality and Social Psychology Bulletin, 25*(7), 873–882. <https://doi.org/10.1177/0146167299025007009>
- Smith, E. R., & Henry, S. (1996). An in-group becomes part of the self: Response time evidence. *Personality and Social Psychology Bulletin, 22*(6), 635–642. <https://doi.org/10.1177/0146167296226008>
- Tajfel, H. (1978). Individuals and groups in social psychology. *The British Journal of Social and Clinical Psychology, 18*, 183–190. <https://doi.org/10.1111/j.2044-8260.1979.tb00324.x>
- Theodoridis, A. G. (2013). Implicit political identity. *PS: Political Science & Politics, 46*(3), 545–549. <https://doi.org/10.1017/S1049096513000681>
- Theodoridis, A. G. (2017). Me, myself, and (I), (D), or (R)? Partisanship and political cognition through the lens of implicit identity. *The Journal of Politics, 79*(4), 1253–1267. <https://doi.org/10.1086/692738>
- Tropp, L. R., & Wright, S. C. (2001). Ingroup identification as the inclusion of ingroup in the self. *Personality and Social Psychology Bulletin, 27*(5), 585–600. <https://doi.org/10.1177/0146167201275007>
- Van Dessel, P., Cummins, J., Hughes, S. J., Kasran, S., Cathelyn, F., & Moran Yorovich, T. (2020). Reflecting on 25 years of research using implicit measures: Recommendations for their future use. *Social Cognition, 38*(SI), S223–S242. <https://doi.org/10.1521/soco.2020.38.supp.s223>
- Verkuyten, M., & Yildiz, A. A. (2007). National (dis)identification and ethnic and religious identity: A study among Turkish-Dutch Muslims. *Personality and Social Psychology Bulletin, 33*(10), 1448–1462. <https://doi.org/10.1177/0146167207304276>

- Wang, X.-Q., Zhu, J.-C., Huo, J.-Y., Liu, M.-F., & Ye, B.-J. (2022). Implicit professional identity: Assessment and relation with explicit professional identity and well-being of pre-service teachers. *European Journal of Teacher Education, 45*(3), 338–355. <https://doi.org/10.1080/02619768.2020.1827385>
- Zou, X., Morris, M. W., & Benet-Martínez, V. (2008). Identity motives and cultural priming: Cultural (dis) identification in assimilative and contrastive responses. *Journal of Experimental Social Psychology, 44*(4), 1151–1159.
- Zou, X., Morris, M. W., & Benez-Martínez, V. (2008). Identity motives and cultural priming: Cultural (dis)identification in assimilative and contrastive responses. *Journal of Experimental Social Psychology, 44*(4), 1151–1159. <https://doi.org/10.1016/j.jesp.2008.02.001>