Journal of Experimental Social Psychology 101 (2022) 104309

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

Social identity switching: How effective is it?☆

Anna K. Zinn [a,*], Aureliu Lavric [a], Mark Levine [b], Miriam Koschate [a,c]

[a] Department of Psychology, University of Exeter, UK
[b] Department of Psychology, Lancaster University, UK
[c] Institute for Data Science and AI, University of Exeter, UK

ARTICLE INFO

Keywords:
Social identity
Identity salience
Identity switches
Multiple identities
Implicit association test
Cross-categorization

ABSTRACT

Psychological theories posit that we frequently switch social identities, yet little is known about the effectiveness of such switches. Our research aims to address this gap in knowledge by determining whether – and at what level of integration into the self-concept – a social identity switch impairs the activation of the currently active identity ("identity activation cost"). Based on the task-switching paradigm used to investigate task-set control, we prompted social identity switches and measured identity salience in a laboratory study using sequences of identity-related Implicit Association Tests (IATs). Pilot 1 ($N = 24$) and Study 1 ($N = 64$) used within-subjects designs with participants completing several social identity switches. The IAT congruency effect was no less robust after identity switches compared to identity repetitions, suggesting that social identity switches were highly effective. Study 2 ($N = 48$) addressed potential differences for switches between identities at different levels of integration into the self. We investigated whether switches between established identities are more effective than switches from a novel to an established identity. While response times showed the predicted trend towards a smaller IAT congruency effect after switching from a novel identity, we found a trend towards the opposite pattern for error rates. The registered study ($N = 144$) assessed these conflicting results with sufficient power and found no significant difference in the effectiveness of switching from novel as compared to established identities. An effect of cross-categorisation in the registered study was likely due to the requirement to learn individual stimuli.

1. Introduction

Research over the last decades suggests a variety of influences of social identities on our daily lives: A salient social identity can affect decision making (Benjamin et al., 2010), the perception of the world around us (Hackel et al., 2018), and how we treat other people based on whether they belong to the same group (Chakravarty & Fonseca, 2017). While theory has suggested that self-categorisation is flexible and context-dependent (e.g., Turner et al., 1994), little is known about the effectiveness of switching between different social identities as an intra-individual process. That is, we know little about an individual's ability to activate a different identity in one's mind from the one that has just been salient.

In contrast, an extensive literature in the domain of cognitive control of attention and performance shows that switching between cognitive tasks incurs a substantial performance cost in the form of longer response times and higher error rates - the "switch cost" (e.g., Rogers & Monsell, 1995). The present research builds on the task switching paradigm and aims to determine whether the psychological process of switching between different social identities is effective or incurs an identity activation cost. Here, identity activation cost refers to an impairment in activating the social identity in one's mind when the social identity changes, resulting in a delayed or weaker activation of the switched-to identity. Further, we address the question of whether any such costs are related to the extent to which social identities are integrated in the self-concept.

1.1. Social identities

Social identities refer to the various group and category memberships that we hold (Tajfel & Turner, 1979). Research over the last 40 years has supported the important role that social identities play in our

☆ This paper has been recommended for acceptance by Dr Ernestine Gordijn.

* Corresponding author at: Department of Psychology, College of Life and Environmental Sciences, University of Exeter, Washington Singer Laboratories, Perry Road, Exeter EX4 4QG, UK.
E-mail address: az274@exeter.ac.uk (A.K. Zinn).

https://doi.org/10.1016/j.jesp.2022.104309
Received 14 April 2021; Received in revised form 13 February 2022; Accepted 25 February 2022
Available online 20 March 2022

lives: They help us to categorise and interact with our social environment in meaningful ways, and provide us with a sense of belonging and social support (Haslam et al., 2018). However, under particular circumstances, social identities can also be the basis for in-group bias and intergroup discrimination (Tajfel et al., 1971; Voci, 2006). Since first proposed as part of Social Identity Theory (SIT; Tajfel & Turner, 1979), the interest in social identities as a concept has grown steadily in different disciplines (Abdelal et al., 2009). Self-Categorisation Theory (SCT) as proposed by Turner et al. (1987) provides the cognitive underpinning to the Social Identity Approach (Abrams & Hogg, 1990). SCT addresses important aspects concerning the flexibility of social identities and factors determining social categorisation. It proposes that, in order to adjust behaviour and cognition to different situations, "self-categorisation is inherently variable, fluid, and context dependent" (Turner et al., 1994, p. 454). The social category that is cognitively activated within a specific social context is referred to as the salient social identity.

There is, however, controversy on how different social identities change in salience and thus the question of how we switch between social identities. Based on the principle of functional antagonism (e.g., Turner & Oakes, 1986), only one social identity can be salient at any given time which would mean that people perform frequent social identity switches in everyday life. This is based on the idea that an individual's personal identity becomes suppressed when one of their group level identities becomes salient and vice versa (Turner et al., 1987). Similarly, to accommodate several social identities within the self, salience needs to shift from one group-level identity to another in a context dependent way. In the more recent literature, the principle of functional antagonism has been criticised as being over-simplified and unable to explain the complexity of groups and behaviours as found in real life (Abrams & Hogg, 2004). For instance, the cross-categorisation approach (Brown & Turner, 1979; Deschamps & Doise, 1978) accounts for more complex categorisation contexts by proposing that two social identities (e.g., age and race) can be simultaneously salient within the same context. This results in an increase in the amount of possible group comparisons. Instead of only being able to distinguish between in-group and out-group, cross-categorisation enables a distinction based on two social identities leading to a double in-group (i.e., a person one shares both group memberships with), double out-group (i.e., a person one shares no group memberships with), or a mixed membership group (i.e., a person that one shares one of the two group memberships with). Moreover, different subgroup identities might also be re-categorised into one superordinate identity (e.g., Gaertner et al., 1989). This could mean, for instance, that rather than switching between one's work identity and parent identity, one might instead identify as a "working parent", hence obviating the need for an identity switch. Notably, neither cross-categorisation nor re-categorisation into a superordinate identity negate the flexible and context-dependent character of social identities. Unless one presumes that *all* of a person's social identities, rather than just two, are salient simultaneously, or can be sub-categorised to form a new social identity (e.g., "working mother"), switches between social identities would still be necessary in many social contexts.

In line with SCT, past studies support the fluidity of social identities by showing that information processing about the self largely responds to shifts in the salient identity rather than being relatively stable (e.g., Onorato & Turner, 2004). This raises the question of how effective such switches are and whether they lead to any impairment in adopting the new identity. To better understand the process of social identity switching, it is also important to account for differences between the identities that are held by an individual. Specifically, the effectiveness of social identity switching might be influenced by the level of integration of the identities in the self-concept (Amiot et al., 2007). The model by Amiot et al. (2007) distinguishes four stages of development of identities and their integration in the self. Throughout the stages, a novel identity develops from (1.) being mainly anticipatory prior to the individual interacting with the new group (also for minimal groups), to (2.) becoming a part of the self but still remaining isolated and highly distinct from already existing social identities, to (3.) becoming increasingly integrated, whereby similarities between novel and older identities are recognised and conflicts resolved, to (4.) becoming well-integrated so that a simultaneous activation of identities becomes possible. For novel identities, Abrams' (1994) proposes that social identity activation might initially need more conscious processing than identities that have become more established in an individual's self-concept, with the latter generally activated in a fluid, "automatic" way. Based on this model, one might expect identity switches to be less effective for novel, less integrated identities, than for established identities that an individual has embedded in their self-concept over time. Accordingly, research into the effectiveness of social identity switches needs to address differences in the integration of social identities in the self-concept to understand whether, and for which social identities, identity activation costs may occur.

Based on the social identity literature discussed above, we aim to address two main gaps in this literature. First, we aim to establish whether social identity switches incur identity activation costs, or whether such switches are highly effective (i.e., activation cost-free). Our second aim is to learn about the conditions that are more likely to lead to such costs. In particular, we will investigate whether the level of integration of identities in the self influences the effectiveness of identity switching. When designing a study that addresses these questions, one must keep in mind that in some contexts two identities might be salient at the same time (e.g., Brown & Turner, 1979; Deschamps & Doise, 1978), or that one identity may have become sub-categorised into the other identity (e.g., Gaertner et al., 1989), as this implies that no switches need to occur between these two identities. In order to better understand how social identity switches may operate, and how these can be measured, we will first review a potentially relevant area of cognitive psychology: task switching.

1.2. Task switching

Switching among simple (or sometimes complex) cognitive tasks is an important part of our everyday lives. The processes that control such switches have been investigated in the laboratory using "task switching" paradigms (e.g., Meiran, 1996; Rogers & Monsell, 1995). In a typical task-switching experiment, participants briefly practice two or more simple tasks such as classifying coloured canonical geometric objects (e.g., squares, circles, triangles) by their colour or by their shape (Lavric et al., 2008; Monsell & Mizon, 2006), or reading a colour name vs. naming the colour of the font - an adaptation of the Stroop paradigm (e.g., Allport & Wylie, 1999). In a within-subject design, participants then complete a large number of trials each lasting a few seconds, with the task either changing from one trial to the next (switch trial) or staying the same as on the previous trial (repeat trial). This allows for a direct comparison of performance (reaction time, or RT, and error rate) on switch trials vs. repeat trials. Such a comparison almost invariably results in a task "switch cost" - slower responses and increased propensity for errors when the task changes (e.g., Meiran, 1996; Rogers & Monsell, 1995). Similar results have been reported for switching between languages for production (speech) in bilinguals (e.g., Graham & Lavric, 2021; Meuter & Allport, 1999). A brief consideration of the possible source(s) of switch costs may inform our subsequent discussion of the potential identity activation costs resulting from switching between social identities.

Several accounts of the sources of task switch costs have been put forward (Monsell, 2003; for more recent reviews see Kiesel et al., 2010 and Monsell, 2015). Two such accounts - which are not mutually incompatible - have received particular scrutiny: "task-set reconfiguration" and "task-set inertia". Task-set reconfiguration refers to the goal-oriented, intentional, change in the configuration (organisation) of the task-set during a switch (Rogers & Monsell, 1995). This can involve the activation of components of the newly-relevant task-set, as well as the

inhibition of components of the previously-relevant task-set. Task-sets may, for instance, require shifts of attention to a different perceptual attribute of the stimulus (e.g., Elchlepp et al., 2017; Longman et al., 2013; Longman et al., 2014; Meiran & Marciano, 2002), or shifts in the relevant set of stimulus-response rules (e.g., Meiran & Marciano, 2002). By analogy, social identities can also differ with regards to goals, attitudes, or attributes (e.g., Amiot et al., 2007). Hence, reconfiguration might also be necessary when switching between social identities, potentially resulting in an identity activation cost. Referring back to the previously discussed model by Amiot et al. (2007), this might be especially the case if the switch includes novel – not yet integrated – identities. Social identities that are less integrated are "highly contextualized and distinct and likely to be associated with distinct thoughts, attitudes, and behaviors" (Amiot et al., 2007, p. 375). When switching involves a novel identity, processes analogous to "task-set reconfiguration" (Rogers & Monsell, 1995) are more likely to be needed because the switch is more likely to require the activation of aspects of an identity that are clearly distinct – maybe even in conflict – with aspects of the previously active identity.

The second account above refers to the "passive" carry-over, or persistence, of the preceding task-set configuration, referred to as 'task-set inertia' (e.g., Allport & Wylie, 1999), which can affect performance on a switch trial by interfering with the newly-relevant task. By analogy, one could expect an already activated social identity to persist and interfere with a newly-activated social identity. This, in turn, might result in ineffective social identity switches with a delayed/weaker activation of the new identity. Conversely, if identities do not show such persistence, then changing which identity is salient is likely to be effective and potentially cost-free.

In order to conduct research on social identity switches, it is important to implement a research design that does not conflate identity switches and other types of switches. An identity switch can be accompanied by a task or a language switch (e.g., from English to Spanish), for instance, when switching from talking to a work colleague to a conversation with a family member. Furthermore, an identity switch can also take place while the task and language remain the same, for example, when a person reads through a work email but a sentence in the email makes their parental identity salient. The task and the language remain the same – writing an email on a specific topic in the same language – but the salient identity changes.

Although, as explained above, one can find clear parallels between the endogenous control of task-set and the control of social identity salience, there is an important difference in the empirical phenomena (and measures) that reflect the cognitive control processes in the two domains. A change of task has a direct (and detrimental) impact on task performance, reflected in longer response times (RTs) and/or a lower response accuracy (higher error rates); this switch-induced worsening of performance is referred to as the task "switch cost" (e.g., Rogers & Monsell, 1995). In contrast, our measure of the effect of switching identity salience is not based on the overall worsening of task performance, but on the reduced difference between the condition where the required behavioural response is congruent vs. incongruent with one's salient identity (the "congruency effect" in the Implicit Association Test; Greenwald et al., 1998, see below). Thus, to acknowledge this important difference between the task switch cost and our measure of the effect of identity switching, we refer to the latter as "identity activation cost" rather than "switch cost".

Notwithstanding the differences between task switches and social identity switches, the task switching research summarised above can form a starting point when conducting research on social identity switches. Based on the task-switching research outlined above, the next section will describe a novel experimental perspective on social identity research which focuses on the effectiveness of social identity switching.

1.3. Switches between social identities

To our knowledge, there are no studies in the social identity/social categorisation literature that have assessed costs of switching identities as compared to maintaining the same identity directly (switch vs. repeat trials), or that have asked whether potential costs of switching depend on the level of integration of identities in the self-concept. The following section will therefore summarise studies that looked at processes related to switching social identities, for instance, shifts between in- and out-group membership and frame switching.

A recent study by Xiao and Van Bavel (2019) examined shifts between in-group and out-group membership in a within-subjects design. Participants were assigned to a red or blue team and asked to complete a sequential priming task. After the first half of the experiment, participants were told they had to change teams due to a computer error, meaning that they had to shift to the initial out-group for the rest of the experiment. Prior to the shift, participants showed a significant out-group bias. After the shift, the RT patterns significantly reversed although the new out-group bias was not significant. This study indicates that, while participants appear to have shifted from in- to out-group, the shift resulted in a reduction of the out-group derogation effect.

These findings indicate that RT patterns are sensitive to changes to an individual's social identity. However, the study by Xiao and Van Bavel (2019) assessed changes in outgroup bias following a change in group allegiance rather than the effect of social identity switches on identity activation. The process of shifting between in- and out-group membership differs from switches in salience between two different social identities that the individual holds.

An example of a within-subjects design that included a switch between cultural identities can be found in the "frame switching" literature. Luna et al. (2008) prompted a switch between participants' cultural identities by asking them to change the language in which they responded during the experiment (English to Spanish, or vice versa). The same IAT was then conducted once in each language to compare the "relative activation strength of culture-specific mental frames" (Luna et al., 2008, p. 279). The findings suggest that biculturals shift mental frames when they change from one language to the other and that such a shift can be prompted and assessed in an experimental design using successive IATs. Specifically, the IAT response patterns showed that – in line with the activated cultural identity – the link between feminine and self-sufficient words as compared to masculine and other-dependent words was stronger in the Spanish IAT compared to the English IAT. This study focuses on the main question of *whether* biculturals (in contrast to e.g., bilingual people who speak two languages but do not hold two different cultural identities) shift mental frames based on the language they use. In contrast, our studies focus on *how effectively* people switch between different identities. The studies by Luna et al. (2008) do not warrant conclusions regarding the effectiveness of social identity switches, because they did not compare the prompted switch to maintaining the same identity (a necessary baseline with which the effect of switch on identity activation should be contrasted), or switches between identities at different levels of integration. However, a very useful methodological lesson from Luna et al. (2008) is that one can use sequences of IATs to elicit identity switches and measure their effects.

Although there have been studies that required shifts between in- and out-groups or between cultural identities, our studies address a gap in research by investigating costs of social identity switches as compared to repeating (staying in) the same identity. To learn more about our ability to switch between social identities, one requires a more controlled study setting which specifically focuses on comparing identity switches to remaining in the same identity – similar to task switching paradigms. First, such a study design should include identity switches and repetitions in order to reach any conclusions on the costs of switching social identities. Second, it is important to avoid conflating social identity switches with other types of switches. For instance, while

we could use language as a prompt to instigate a social identity switch in bicultural individuals, any costs found in this study design could not be clearly attributed to the social identity switch as participants would also switch languages. Furthermore, a language switch does not necessarily require an identity switch: Not everyone that speaks German self-categorises as 'German', for example. The present research therefore examines patterns of RTs and errors during social identity switches to test for the presence of identity activation costs of switching social identities. To take a first step towards establishing whether social identity switches lead to any identity activation costs, Pilot 1 and Study 1 focus on switches between well-established, neutral to positively valued identities as such identities represent the majority of social identities (e.g., Deaux et al., 1995) and have been the initial focus of SIT theorising (Tajfel & Turner, 1979). The studies directly compare identity switches with remaining in the same social identity (identity repetitions). Study 2 and the pre-registered study will extend this by examining whether the effect differs when the switch occurs away from a novel or well-established identity, respectively.

2. Study 1: identity activation costs between well-established, positively-valenced identities

In order to assess identity activation costs, two important aims of the first study[1] were to: (1) establish a study design and (2) identify a measurement method for assessing several social identity switches. The previously summarised theory posits that people engage in relatively rapid switches between social identities that can be prompted by changes in the environment (e.g. Turner et al., 1994). Hence, our initial study on identity activation costs was conducted in a laboratory setting to keep external prompts that might lead to influences on the salience of social identities to a minimum. One way to assess social identity salience implicitly is to infer the social identity from behaviours or response patterns caused, or influenced, by the salient identity (Koschate et al., 2021). One such typical response pattern is the positivity bias – favouring the in-group over the outgroup (in-group favoritism as shown e.g., by Otten & Moskowitz, 2000).

The present study uses the Implicit Association Test (IAT; Greenwald et al., 1998) as a means of prompting relatively rapid social identity switches and to determine the salient identity by measuring the resulting positivity bias towards the in-group (e.g., Otten & Moskowitz, 2000) based on the IAT congruency effect. The primary reason for using the IAT in our study is that it can be reconfigured so that the task itself (i.e., categorising in- and out-group stimuli for the target category) makes a social identity directly salient through its instructions and task requirements. This will ensure that the task performed by participants is directly related to the relevant social identity and therefore will likely keep the identity salient. The IAT is also a subtle and relatively rapid assessment. Previous studies (Luna et al., 2008) have shown that the IAT can be used for prompting different identities in a within-subject design. A study by Nosek et al. (2002) has also shown that several different IATs presented in a sequence yielded scores that were highly consistent with scores from a dataset where each participant only completed one IAT. This further shows that IATs can be used to obtain repeated measures of social identity switching in a within-subject design. In what follows, we detail how we adapted the IAT for our study.

The IAT has been implemented in numerous studies across several fields of psychology (Hofmann et al., 2005). The typical IAT "congruency" effect refers to participants making faster key-press responses on trials in which the target and attribute items that share a key are expected to be highly associated ("congruent" trials) compared to trials on which the associated targets and attributes are responded to with different keys ("incongruent" trials, Greenwald et al., 1998). Our IATs are based on the classic IAT (Greenwald et al., 1998), with in-group faces mapped onto the same response as positive words and out-group faces mapped onto the same response as negative words on congruent trials. Conversely, on incongruent trials, in-group faces and negative words are mapped onto the same responses, as are out-group faces and positive words (see Fig. 1).

Importantly, the overall increase in error rates and RTs, that is, the main effect of switching from one IAT to another (e.g., race IAT to age IAT) will mostly be driven by the effects of a task switch (switching between the task of categorising a face by race to categorising it by age as compared to continuing with the same categorisation task; see Fig. 1). As explained in the section on the task-switching literature, a study assessing an identity activation cost must avoid conflating social identity switches with task switches. Therefore, we will focus on whether the congruency effect (the positivity bias towards the in-group) is affected by switching identities rather than the overall RTs and error rates. Hence, the statistical effect we need to focus on corresponds to an interaction effect between congruency (congruent vs. incongruent target/attribute combination) and switching (switch vs repeat trials) rather than on the statistical main effect of switching from one IAT to another. In other words, the analysis of the congruency by switch interaction unconfounds the effect of a change of social identity from the effect of a change of task.

IAT studies examining intergroup bias typically calculate a difference measure between congruent and incongruent trials - the D-Score (Greenwald et al., 2003). Given that we are not interested in the interpretation of bias, but in inferring identity activation costs from the congruency effect, we will instead compare congruent and incongruent trials and analyse RTs and error rates separately as is typically done in task switching research. The D-Score also includes a factor that scales the variance of interest by the nuisance variance - this can be useful when the primary aim is to correlate the IAT congruency measure with other, non-IAT measures. Instead, our aim is to subject the congruency effect itself to statistical significance tests (ANOVAs) which already scale by the nuisance IAT variance. As a result, we see no reason to scale by nuisance variance twice - while computing the D-Score and as part of the significance testing.

Study 1 aims to answer the question of whether there is an identity activation cost (an impairment caused by delayed, or weakened, activation of the new identity) associated with having to switch to another identity as compared to remaining in the same social identity. Based on task switching research (e.g., Meiran, 1996; Rogers & Monsell, 1995), we would expect switching between identities to be demanding (effortful). Indeed, identity activation costs might arise from a delay in shifting to attributes of the new identity (similar to "task-set reconfiguration"; Rogers & Monsell, 1995) or from the persistence of the previous identity (similar to "task-set inertia"; e.g., Allport & Wylie, 1999). Assuming that the in-group positivity bias in the IAT depends on in-group salience, an impairment of activation should result in reduced bias compared to when the same identity is repeated – leading to Hypothesis 1a. Appendix A includes graphs showing the expected result patterns in line with each of the following hypotheses.

H1a. The congruency effect is reduced for switches between different salient identities compared to repeating the same salient identity, indicating higher identity activation costs during a switch.

Conversely, SCT (Turner et al., 1987) describes the process of self-categorisation as both "inherently flexible and context-dependent" (Turner et al., 1994, p. 454). Based on this definition, it may be assumed that switching between social identities is highly effective and does not result in any impairment (be it delayed or weaker identity activation). Therefore, if social identity switches are indeed highly effective, we would expect the bias towards favouring the newly activated in-group to be (largely) unaffected by switches. This leads to the following contrasting Null Hypothesis:

[1] OSF anonymous pre-registration link for Study 1: https://osf.io/4v3gu/?view_only=eb19c77240db4b65a84e6a811ebe38f9

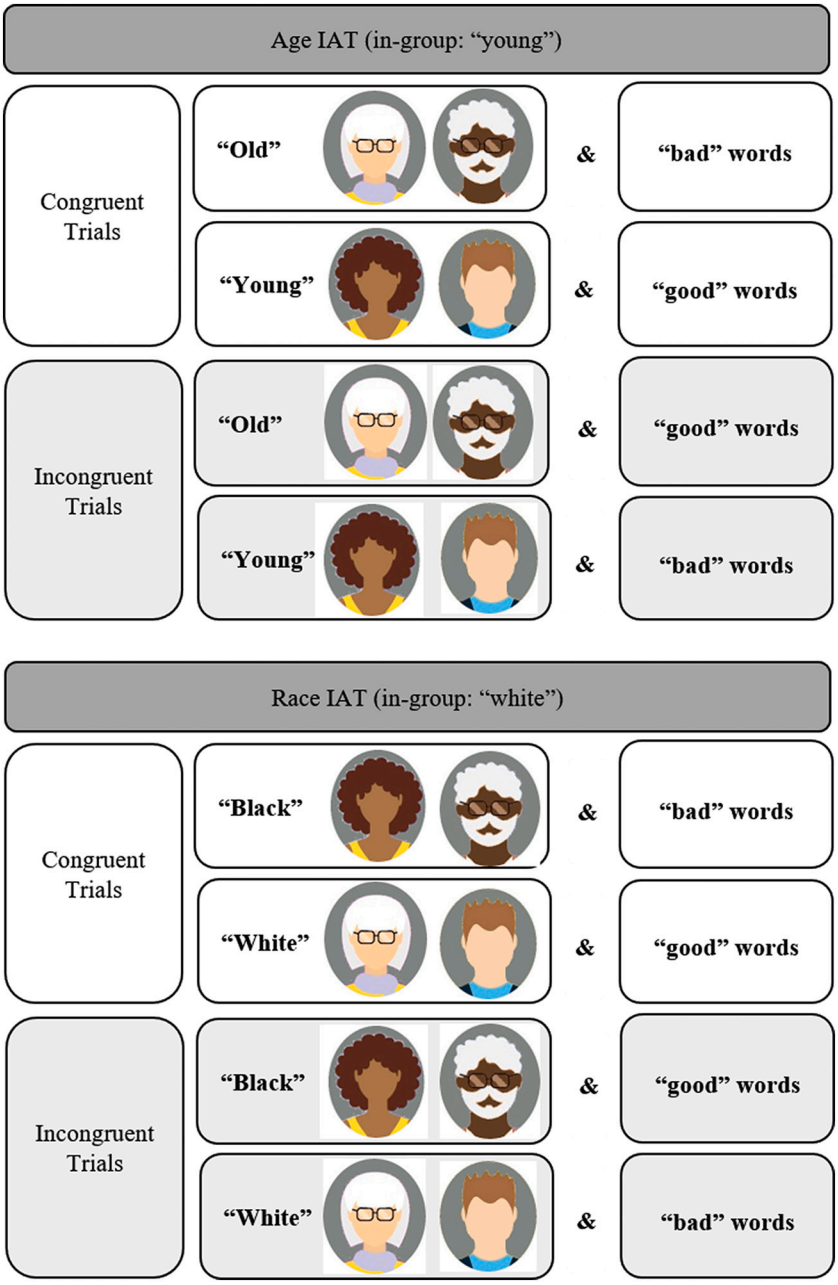

Fig. 1. Congruent vs. incongruent trials depending on type of IAT.
Note. This figure shows example avatars (©iconsy via canva.com) for illustration – in the study we used actual face images from pre-tested face databases (Bainbridge et al., 2013; Minear & Park, 2004); the amount of male vs. female faces was balanced for each condition.

H1b. (Null Hypothesis): The congruency effect is equivalent in size for switches between different salient identities compared to repeating the same salient identity, indicating no effect of switching on identity activation.

It may also be that the detrimental effects of an identity switch are confined to a short period of time following the switch. In task switching research (e.g., Meiran, 1996; Rogers & Monsell, 1995), individual trials typically last seconds rather than minutes (as in IATs). Therefore, based on the task switching literature and in conjunction with H1a, it can be expected that social identity activation costs are larger at the beginning of a switch-related IAT.

H2. Switching between identities will affect the congruency effect more strongly in the first part of the IAT than in the second part of the IAT.

Finally, we accounted for the possibility that instead of switching between the two identities, identities might overlap or become salient concurrently. Based on cross-categorisation (Brown & Turner, 1979), we tested whether more than one social identity was salient at the same time. In our study participant's in-groups were "young" and "White". As shown in Fig. 2, the images in the study held either attributes of both in-groups (here: "young, White person"), both out-groups (here: "old, Black person"), or a mix between in- and out-group attributes (here: "young, Black person"; "old, White person"). This allowed us to compare patterns of results between those different types of stimuli.

Cross-categorisation states that double in-groups should be evaluated more positively than double out-groups, while mixed-membership groups are expected to fall somewhere between those two groups (see Crisp, 2010). In an IAT, this would become evident in double in-group faces being strongly associated with positive words and double out-

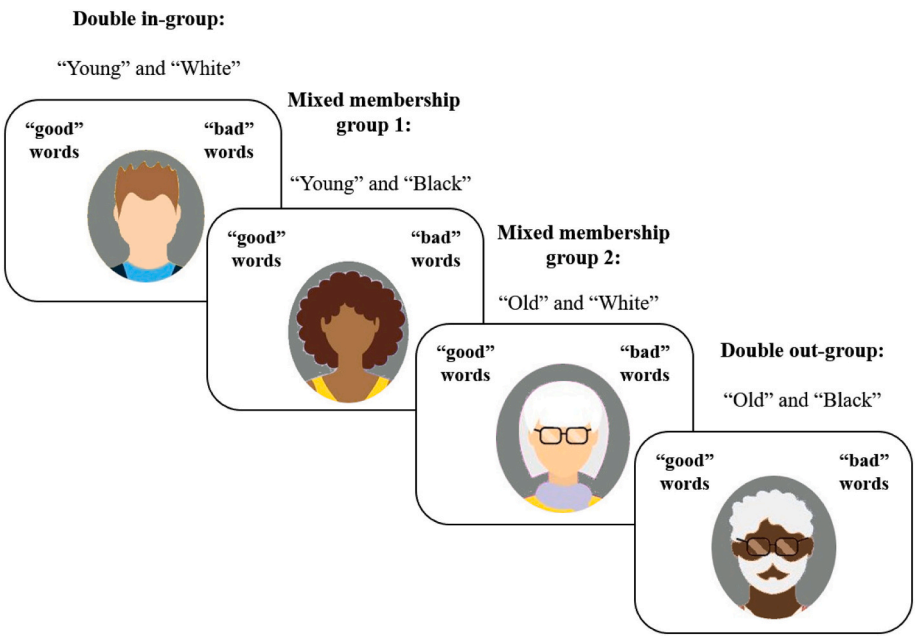

Fig. 2. Different types of face stimuli based on cross-categorisation hypothesis.
Note. This figure shows example avatars (©iconsy via canva.com) for illustration – in the study we used actual face images from pre-tested face databases (Bainbridge et al., 2013; Minear & Park, 2004).

group faces with negative words, resulting in an overall larger congruency effect for those stimuli. Mixed membership stimuli, in contrast, would not be expected to show such a strong association with negative or positive words, resulting in a smaller congruency effect. Conversely, if the manipulated identities are not concurrently salient (only the identity relevant in the context of the IAT being performed is salient), we would not expect to find differences between mixed membership and double in−/out-group stimuli.

Based on the above reasoning, if both identities remain salient concurrently during a switch (e.g., participants think of themselves as "young" and "White"), we would expect a smaller congruency effect for mixed membership stimuli compared to double in-group or double out-group stimuli.

H3. Switching between salient identities is expected to reduce the congruency effect more strongly for trials with mixed membership group stimuli as compared to double in- or double out-group stimuli.

2.1. Methods

2.1.1. Participants and design

A total of $N = 64$ participants[2] (84% female, 14% male, and 2% gender fluid) were included in Study 1 in line with sample size calculations based on Brysbaert and Stevens (2018).[3] Due to the nature of the social identities (age, race) tested, we recruited participants that were young (18–25 years old; $M = 19.92$, $SD = 1.61$) and considered themselves to be White. Participants had to be proficient English speakers and have normal or corrected vision to be able to complete the IAT. The majority of participants (72%) were British and had English as their first language (78%). Most (84%) were undergraduate full-time students, with 61% studying Psychology. Finally, 83% of participants were right-handed and 14% left-handed (3% missing responses).

The main study has a 2 (congruency: congruent vs. incongruent trials) x 2 (switch: switch vs. repeat trials) x 2 (social identity salience: age vs. race) within-subjects design[4] with RTs and error rates as the dependent variables. Pilot 1 and Study 1 received approval from the departmental ethics committee at the University of [blinded for review].[5]

2.1.2. Materials and measures

The study uses five successive IATs to study effects of identity switching: One starting social identity IAT plus four IATs to measure potential identity activation costs of switching versus remaining in a salient identity.

2.1.2.1. Implicit association test. Although the IATs used in this study are based closely on best practice recommendations (e.g., Greenwald et al., 2003; Greenwald et al., 2020) the IAT had to be adapted to the specific within-subjects design. Compared with the standard IAT recommended by Greenwald et al. (2020), three main changes were made:

1) To allow for more rapid switches between identities, the IATs did not include a separate practice session prior to the first part of each IAT. Participants were simply informed about the social category (age or race) that target stimuli needed to be categorised by in the next IAT. Midway through each IAT, a short practice of the face stimuli was provided once the target categories changed finger allocations.
2) No category labels (e.g., Black/White, good/bad) were displayed on the left- and right-hand side of the screen in order to avoid prolonging RTs (and increasing the nuisance variance) through occasional eye movements to read the category labels. It is possible that

[2] We tested until we reached the minimum required sample size of 64 participants. In total 70 participants completed the study of which 6 did not meet the study inclusion criteria. The data of those 6 participants was therefore not included in any analysis or data summary. The data was only analysed once data collection had been completed.

[3] For sample size calculations please see supplementary materials S1.1

[4] Pilot 1 followed the same design as Study 1. However, due to the smaller sample size, Pilot 1 focused on the main effect of congruency as well as on identifying outliers in the word and face stimuli.

[5] We disclose all measures, manipulations, and exclusions in this study. We report how we determined the final sample size and whether data collection was continued after data analysis.

eye movements are more common in incongruent trials which would inflate incongruent RTs – and therefore the congruency effect.
3) Since we are interested in identity switching that might occur early on during an IAT, the same number of trials (64) were used in each part of the IAT – resulting in 2 main sections of each IAT (one containing congruent trials and one containing incongruent trials) rather than the 4 analysed sections in the "standard" IAT.

The modified IATs, including all target and attribute stimuli, were piloted in a within-participants design prior to the main study, with a different set of participants ($N = 24$). The results from Pilot 1 (see Appendix B) confirmed the expected robust (and significant) main effect of congruency for both identities. These findings suggest that our adapted IATs run successively are sensitive to in-group bias and can therefore be used to assess identity switching.

2.1.2.2. Social identities. In the present study, we used age and race as social identities. We chose these two identities as in-group/out-group members can be visually identified from the face stimuli in the IAT, and the same pictures can be used in the race IAT and the age IAT. Based on the face stimuli we can also distinguish between double in-group (young, white faces), double out-group (old, Black faces) and mixed membership groups (young, Black faces; old, White faces) for our particular sample (young, White participants).

The salient identity was activated by instructions for the IAT and error messages in case of incorrect categorisation. Additionally, a short prompt appeared on the screen telling participants that either an age or a race IAT was about to start.

2.1.2.3. Stimuli. The study included four main stimulus sets, each consisting of faces and words (one for each of the four main IATs), plus a set for the starting IAT. The order in which the four stimulus sets used in the main IATs were presented was balanced over participants. Each stimulus set consisted of eight different positively-valenced ('good'), eight negatively-valenced ('bad') words (based on valence ratings by Warriner et al., 2013), and four photographs of faces of different individuals per combination of age and race (young White, young Black, old White, old Black). Half of the faces in each combination of age and race were female and the other half male. Face stimuli were taken from pre-tested databases (Bainbridge et al., 2013; Minear & Park, 2004). We ensured that the background and cropping was similar across face photographs. Based on the age ratings collected by the authors of the databases (Bainbridge et al., 2013; Minear & Park, 2004), Black and White faces we selected for the "young" category were perceived to be, on average, of similar age ($M_{Black} = 20.88$, $SD = 1.89$; $M_{White} = 21.00$, $SD = 1.67$), and so were the faces we selected for the "old" category" ($M_{Black} = 59.81$, $SD = 11.55$; $M_{White} = 66.69$, $SD = 7.31$).

All stimuli were tested in Pilot 1. The face stimuli had similar accuracy scores ($M = 94.3\%$, $SD = 1.95$) with only one stimulus showing accuracy scores that were > 3 SDs below the mean. For the main study, this picture was replaced with a photo from the previous practice IAT that showed a higher accuracy. The mean RT was 827 ms ($SD = 42.66$) for faces and 980 ms ($SD = 64.11$) for words, without any obvious outliers for either type of stimulus.[6] The average accuracy for the words was 93.6% ($SD = 2.67$); one word had a value >3 *SD*s below this value and was replaced in the main study.

2.1.2.4. Congruency. Each IAT consisted of 64 congruent trials (in-group pictures and positive words mapped onto the same finger) and 64 incongruent trials (in-group pictures and negative words mapped onto the same finger). Whether IATs started with congruent or incongruent trials was balanced between and within participants; The allocation of the "good" and "bad" word categories to response fingers was balanced over participants (see Appendix C).

2.1.2.5. Switching. Participants completed one switch and one repeat IAT for each identity. Therefore, "repeat trials" refer to trials in an IAT that requires the same categorisation as the previous IAT (e.g., two age IATs). "Switch trials" refer to the trials in an IAT with a different categorisation than the previous IAT (e.g., participants had to switch from an age to a race IAT). The order of switch vs. repeat IATs and race vs. age IATs was balanced over participants (see Appendix C).

2.1.2.6. Demographics and identification questionnaire. [7] The paper-based questionnaire included demographic questions as well as items assessing the strength of identification with the age and race identity based on Doosje et al. (1995) and Haslam et al. (1999), interaction with the in-group and out-group, and open text questions asking whether participants knew what the study was about, tried to influence their results and how easy they perceived it to think of themselves as "young" and as "White". Strength of identification, interaction with in- and out-group, and the open text questions were included to find out more about potential issues with the design in case the preconditions were not met, but these are not included in the analysis presented here.

Although Pilot 1 also included a self-report measure of identity salience (based on Verkuyten & Hagendoorn, 1998) after each IAT, this measure was removed for Study 1 to reduce the time between switches. In Pilot 1, the questionnaire was completed at the start of the study, but it was moved to the end of the main study to avoid making identities salient before the start of the IATs.

2.2. Procedure

The main part of the study including instructions, tasks and measurements were conducted on a computer using *E*-prime 2.0 (Psychology Software Tools, Pittsburgh, PA; see Schneider et al., 2002), with the paper-based questionnaire being completed after the main experiment. The experimenter had only minimal interactions with participants to greet them, discuss any questions regarding informed consent, prompt them to fill out the questionnaire, and to debrief and reimburse them. The study took approximately 20–30 min.

After receiving information about the study and signing a consent form, participants were asked to use a computer in a soundproof cubicle for the main part of the study. It consisted of IAT practice trials, followed by the full starting IAT and the four main IATs. Due to the within-subjects design of the study, all participants received the same treatment (one switch and one repeat IAT per identity) but in a counterbalanced order that the experimenter was not aware of. After finishing the IAT part of the study, participants were asked to fill in the demographics and identification questionnaires and were then fully debriefed by the experimenter. Participants received either course credits or a financial reimbursement of £5 for their participation.

2.3. Results

RTs under 200 ms, practice trials and the starting IAT were excluded from the analysis. Only trials with correct responses were included in the

[6] Due to an error (detected after data analysis was completed) one face stimulus was included twice in the experiment – in two different stimulus sets. However, because this stimulus was not associated with outlier responses, and since the order of the stimuli sets was balanced over participants this is unlikely to have affected the results

[7] For the full questionnaire and a summary of descriptive statistics not included in this report please see supplementary materials S2

RT analysis, as is typically done in task-switching studies. Since all 64 participants were within 3 *SD*s of the mean for both RT and error rate, they were all included in the analysis.[8]

2.3.1. Precondition and Hypothesis 1 A and B – effectiveness of switching

Before testing our hypotheses, we tested whether the prompted identities were indeed salient during the study, that is, whether participants showed the expected bias towards evaluating the in-group more favourably than the out-group. We did so by testing whether RTs were faster for congruent compared with incongruent trials, including tests for the effect of congruency for each of the two identities.

A 2 (congruency) x 2 (switching) x 2 (identity) repeated-measures ANOVA showed the expected significant main effect of congruency (see Table 1 for significance tests). Participants had slower responses in incongruent (M_{Age} = 927; M_{Race} = 882) compared to congruent trials (M_{Age} = 811; M_{Race} = 801). We also found a significant interaction between congruency and identity, and a main effect of identity with participants responding overall faster in race (M = 841) compared to age IATs (M = 869), hence we conducted separate follow-up tests looking at the congruency effect for the two identities separately. The results showed a significant congruency effect for both age ($F(1,63) = 87.95$, $p < .001$, $\eta_p^2 = 0.583$) and race ($F(1,63) = 43.3$, $p < .001$, $\eta_p^2 = 0.407$).

For the error rates we also found the expected significant main effect of congruency (see Table 2) with higher error rates in incongruent (M = 7.3%) compared to congruent trials (M = 4.5%). There was no significant main effect of identity or significant interaction between identity and congruency. These results for both RTs and error rates fulfil the preconditions for hypothesis testing - that each of the two identities became salient when the IAT categorisation made it relevant.

To test whether identity switches affect the congruency effect (Hypothesis 1 A and B), we focused on the interaction between congruency and switching (see Fig. 3). The findings for RT as the dependent variable show no significant interaction (see Table 1), thus providing no support for hypothesis 1a. The findings for error rates showed a significant interaction between congruency and switching (see Table 2). However, the pattern of findings did not support Hypothesis 1a either, because the congruency effect was larger in switch trials (M_{Con} = 4.3%; M_{Inc} = 7.8%; M_{Diff} = 3.5%, SE = 0.51)[9] compared to repeat trials (M_{Con} = 4.8%; M_{Inc} = 6.8%; M_{Diff} = 2.0%, SE = 0.37). Switching did not reduce the congruency effect in RTs or error rates. This suggests that the switches between salient social identities were very effective, in line with the null hypothesis b.

Finally, the main effect of switching for RTs is in line with findings from the task switching literature, with slower responses in switch (M = 871) compared to repeat IATs (M = 839), indicating that the effect of a change in salient identity showed at least some task-switch costs during the IAT immediately following a switch. The main effect of switching was non-significant for error rates.

Table 1
2 (congruency) x 2 (switching) x 2 (identity) ANOVA results for RTs.

Effect	*F*	*df*	*p*	η_p^2
Congruency	103.46	1, 63	<0.001	0.622
Switching	31.68	1, 63	<0.001	0.335
Identity	18.00	1, 63	<0.001	0.222
Congruency x Switching	0.96	1, 63	0.331	0.015
Congruency x Identity	5.22	1, 63	0.026	0.077
Switching x Identity	1.32	1, 63	0.256	0.020
Congruency x Switching x Identity	0.03	1, 63	0.859	0.001

Table 2
2 (congruency) x 2 (switching) x 2 (identity) ANOVA results for error rates.

Effect	*F*	*df*	*p*	η_p^2
Congruency	59.77	1, 63	<0.001	0.487
Switching	0.97	1, 63	0.328	0.015
Identity	0.91	1, 63	0.345	0.014
Congruency x Switching	7.92	1, 63	0.007	0.112
Congruency x Identity	2.38	1, 63	0.128	0.036
Switching x Identity	0.27	1, 63	0.605	0.004
Congruency x Switching x Identity	0.55	1, 63	0.463	0.009

Hypothesis 2. Early effect of switching.

A 2 (congruency) x 2 (switching) x 2 (time: first vs. second half of each IAT) ANOVA was conducted to test whether switching between identities affects the congruency effect more strongly in the first part than in the second part of the task. Hypothesis 2 – and thus an impairment in activating a new identity - could not be supported, since we found no significant interaction between congruency, switching and time for RTs (see Table 3) or error rates (see Table 4).

There was a significant main effect of time, with overall slower responses in the first (M = 873) compared to the second part (M = 837) of IATs. This effect was also significant for error rates but with lower error rates in the first (M = 5.5%) compared to the second part (M = 6.3%) of IATs. There was a significant interaction between switching and time with a larger RT difference between repeat and switch trials in the first (M_{Switch} = 903; M_{Repeat} = 844; M_{Diff} = 59.17, SE = 10.29) than the second half (M_{Switch} = 839; M_{Repeat} = 834; M_{Diff} = 5.81, SE = 8.01) of an IAT. This interaction was also significant for error rates - there too the switch-repeat difference was bigger in the first (M_{Switch} = 6.0%; M_{Repeat} = 4.9%; M_{Diff} = 1.1%, SE = 0.41) than the second half (M_{Switch} = 6.1%; M_{Repeat} = 6.6%; M_{Diff} = −0.6%, SE = 0.43). These two-way interactions indicate that the task-switch cost (not to be confused with the identity activation cost) was larger - or only present - in the first half of an IAT.

Hypothesis 3. Cross-Categorisation.

If both identities were salient concurrently during a switch, we would expect a larger congruency effect for face stimuli that present either double in-group attributes (young, White faces) or double out-group attributes (old, Black faces) compared to mixed membership group stimuli (young, Black faces; old, White faces) in switch trials. The 2 (congruency) x 2 (switching) x 2 (type of stimulus: mixed membership group vs. double in-or out-group stimuli) ANOVA did not show this pattern. There was no significant interaction between congruency, switching and type of stimulus for RTs ($F(1,63) = 0.03$, p = .858, η_p^2 = 0.001) or error rates ($F(1,63) = 0.08$, p = .782, $\eta_p^2 = 0.001$). This means that Hypothesis 3 - a reduced congruency effect for mixed membership group stimuli after switching - is not supported.

The two-way interaction between congruency and type of stimulus tests whether both identities were active simultaneously throughout the whole experiment (rather than just after switches). This interaction was marginally significant for RTs ($F(1,63) = 3.95$, p = .051, η_p^2 = 0.059). Irrespective of whether it was a repeat or switch trial, mixed membership group stimuli overall showed a larger (rather than the expected smaller) congruency effect (M_{con} = 739; M_{inc} = 852; M_{Diff} = 113, SE = 12.53) compared with double in-group and out-group stimuli (M_{con} = 746; M_{inc} = 840; M_{Diff} = 94, SE = 12.47). The error rate showed the reverse pattern for the significant congruency by type of stimulus interaction, with $F(1,63) = 9.59$, p = .003, η_p^2 = 0.132 (mixed membership group stimuli: M_{con} = 5.8%; M_{inc} = 6.1%; M_{Diff} = 0.3%, SE = 0.61; double in-group and out-group stimuli: M_{con} = 3.5%; M_{inc} = 6.4%; M_{Diff} = 2.9%, SE = 0.68). Taken together, this pattern of two-way interactions is ambiguous and does not allow us to draw conclusions with regards to cross-categorisation (H3).

[8] Since the present studies have within-subject designs, the SEs of individual means are not informative regarding the variability of the contrasts of interest (e.g., the difference between congruent and incongruent trials). Therefore, they are not provided in the text or in the figures. We report, however, the SEs of the contrasts.

[9] M_{Con} = Mean for congruent trials; M_{Inc} = Mean for incongruent trials

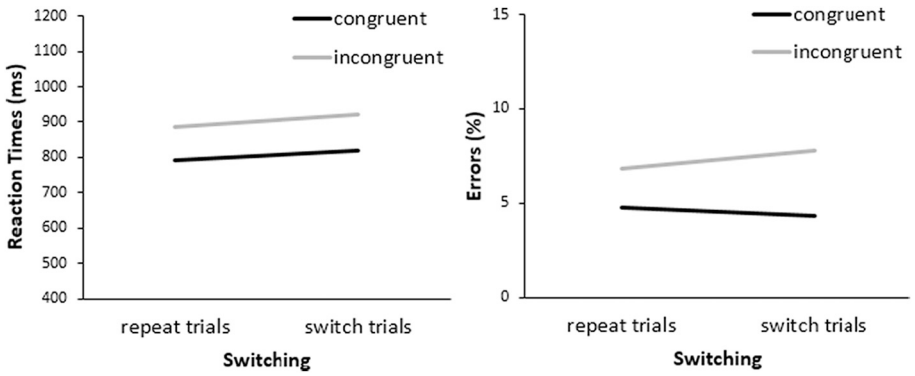

Fig. 3. RTs and error rate as a function of congruency and switching.

Table 3
2 (congruency) x 2 (switching) x 2 (time) ANOVA results for RTs.

Effect	*F*	*df*	*p*	η_p^2
Congruency	103.46	1, 63	<0.001	0.622
Switching	31.68	1, 63	<0.001	0.335
Time	31.27	1, 63	<0.001	0.332
Congruency x Switching	0.96	1, 63	0.331	0.015
Congruency x Time	0.44	1, 63	0.510	0.007
Switching x Time	13.76	1, 63	<0.001	0.179
Congruency x Switching x Time	1.24	1, 63	0.269	0.019

Table 4
2 (congruency) x 2 (switching) x 2 (time) ANOVA results for error rates.

Effect	*F*	*df*	*p*	η_p^2
Congruency	59.77	1, 63	<0.001	0.487
Switching	0.97	1, 63	0.328	0.015
Time	7.49	1, 63	0.008	0.106
Congruency x Switching	7.92	1, 63	0.007	0.112
Congruency x Time	1.71	1, 63	0.196	0.026
Switching x Time	6.83	1, 63	0.011	0.098
Congruency x Switching x Time	2.13	1, 63	0.150	0.033

2.3.2. *Exploratory analysis*

2.3.2.1. "Faking" IAT scores. According to research on "faking" IAT scores, participants may try to reduce their IAT score once they gained an understanding of the task in order to appear less biased. They may do so by responding more slowly during the parts of the IAT containing congruent trials and responding more quickly during the parts containing incongruent trials (e.g., Röhner et al., 2013). In these studies, however, participants were not successful in changing their IAT score by increasing their error rates. Therefore, this exploratory analysis will focus on RTs only.

We analysed the RTs for congruent and incongruent trials of participants completing a repeat vs. a switch IAT after the initial practice IAT (see Table 5). If the second IAT was a repeat IAT, there was no significant RT difference for both congruent and incongruent trials between the first and second IAT. If the second IAT was a switch IAT, participants showed significantly slower reactions for congruent trials in the second IAT compared to the first. The pattern of results - slowing on congruent trials in the second IAT - shows that participants might have tried to decrease their IAT scores to appear less biased.

2.4. *Discussion*

The results of Pilot 1 and Study 1 indicate that the IAT is a method that can be used to assess frequent social identity switches. The findings show that switches are comparatively effective even when relatively rapid switches are required: participants showed the expected congruency effect in line with the newly prompted identity and switching did not seem to impair the activation of a different identity. The findings contribute to the existing literature by providing the first experimental support to theories that suggest that identity switches are a fluid (e.g., Turner et al., 1994) and possibly relatively automatic, in that they may not require intentional control (e.g., Abrams, 1994). We also tested the possibility that more than one identity might be salient at a given point (Brown & Turner, 1979) which would obviate the need to switch identities. Study 1 did not find support for co-activation of identities on switch trials. However, we found an ambiguous pattern of results concerning the co-activation of identities throughout the study (averaging over repeat and switch trials). While this pattern at the very least does not suggest a strong co-activation of the identities, one of the main aims of the planned registered report is to shed more light on the possibility of the co-activation of identities and its interaction with rapid identity switches (see below).

We also considered the possibility that participants might try to fake their IAT scores to avoid socially undesirable responses (when out-group faces and negative words are mapped onto the same responses). Our exploratory analysis found some indication (numerically), that participants slowed down their responses to congruent trials, suggesting that they might have tried to fake lower scores, significantly only for switch IATs. These findings are in line with several studies that have reported people successfully "faking" IAT scores (Fiedler & Bluemke, 2005; Lai et al., 2016). In the present study design, participants might have been able to reduce their congruency effect once they had time to gain an understanding of the task. A reduction of the congruency effect in repeat trials could have masked the interaction of congruency and switching. However, this reduction was numerically small and non-significant when the same IAT was repeated and thus it is unlikely to have masked social identity activation costs. As we explain later, Study 2 and the registered study compare the effectiveness of switches from different types of identity rather than repeat vs. switch IATs, hence avoiding potential issues of participants trying to influence their IAT scores in repeat IATs.

Another issue that we address in Study 2 and the registered study is that the current paradigm does not measure self-categorisation directly, but rather, via an implicit bias that results from self-categorisation. Self-categorisation is the process underlying phenomena such as normative behavior and - as partially assessed by the congruency effect - stereotyping and positive in-group bias (Hogg & Terry, 2000). Although we clearly observed in-group favoritism in the form of the congruency effect in the present study, and made this a precondition for our analyses, this precondition might not be met for other identities or participant groups. In particular, since the congruency effect for the type of IAT used in Study 1 relies on valence matches (in-group with positive words and out-group with negative words) this would not be optimal when assessing social identity switches involving negatively valenced or stigmatised

Table 5
RTs in congruent and incongruent trials in first switch vs. first repeat IAT.

		First IAT *M*	Second IAT *M*	M_{Diff}	*SE*	*t*	*df*	*p*
Switch ($n = 32$)	congruent trials	797	863	66.65	23.12	−2.88	31	0.007
	incongruent trials	1020	977	−42.80	28.62	1.50	31	0.145
Repeat ($n = 32$)	congruent trials	762	781	18.35	14.16	−1.30	31	0.205
	incongruent trials	918	888	−30.07	20.97	1.43	31	0.162

groups. For instance, older people do not typically show the expected in-group favoritism in the IAT, but, quite the reverse, a relatively strong positive implicit bias towards young faces (Nosek et al., 2002). To ensure that in our second study we measure self-categorisation more directly (rather than indirectly through in-group favoritism), we decided to use identification IATs (e.g., Pinter & Greenwald, 2010). Identification IATs assess the strength of association between in-group and out-group stimuli and self- vs. other-related words (Pinter & Greenwald, 2010) and results are therefore more clearly attributable to changes in self-categorisation. Moreover, the congruency effect can no longer be driven by differences in valence.

Finally, while the present study supports the account that social identity switches are rapid and effective, it focuses on two specific identities (age and race), and this conclusion might not apply to other identities. Hence, it will be one of the main aims of Study 2 and the registered study to address potential differences in the effectiveness of switches between different types of identities. Specifically, those studies will assess whether there are differences in one's ability to switch between relatively well-established, positively valenced identities (as tested in Study 1) compared to novel identities. This comparison can help us understand challenges involved in forming new group memberships and integrating them into our everyday lives.

3. Study 2

Our first study suggests that the frequent social identity switches one performs in everyday lives are highly effective - at least for neutral to positively valenced identities that participants have already held for a relatively long time. Forming new group memberships is an important and necessary aspect in our lives, with various benefits linked to having multiple meaningful and positive group memberships (for an overview see e.g., Haslam et al., 2009; Haslam et al., 2018). The question that arises is whether social identity switches are equally effective for novel identities, or whether there is an initial phase – of integrating the new social identity as a part of ourselves and our other social identities (Amiot et al., 2007) - during which switches are costly. Hence, Study 2[10] and the proposed registered study will focus on the comparison of switching away from a novel social identity vs. switching away from an already established social identity to another established identity. Since individuals have not had the chance to integrate a completely novel identity in the self-concept, but have had ample time to integrate the "established" identities, a comparison of activation costs lets us directly assess whether integration into the self-concept supports effective identity switches.

Two theoretical models support the idea that identity integration may affect the effectiveness of identity switches: Social self-regulation theory (Abrams, 1994) and the cognitive-developmental model of social identity integration (Amiot et al., 2007).

Social self-regulation theory (Abrams, 1994) combines social identity and self-awareness approaches, and distinguishes between unconscious and more controlled processes. It suggests that making established identities salient is an automatic and non-conscious process. This is in line with our findings above that indicate that social identity switches are effective and relatively fast. However, this may not be the case for novel identities. Abrams (1994) proposed that if a specific group is made salient for the first time (as it is, for instance, in a minimal group paradigm) "the perceptual/interpretive (self-defining) process may require conscious processing" (Abrams, 1994, p.478). This effort decreases, and the process becomes more automated with repeated exposure to the new social identity.

Similarly, the cognitive-developmental model of identity integration by Amiot et al. (2007) suggests that a novel, less integrated social identity may impair an identity switch. The model distinguishes between early stages of identity integration where a novel social identity is distinct from already existing social identities, and later stages where the identities are highly integrated. A novel identity that is not yet integrated in the self-concept requires distinct, potentially conflicting, components of the identities involved in the switch to be activated. This activation likely involves effortful cognitive control processes, akin to "task-set reconfiguration" in the task switching literature (Rogers & Monsell, 1995). Conversely, a switch between two highly integrated identities that share components such as goals and attitudes should require less cognitive control processes, and consequently result in lower identity activation costs than a switch from a novel identity.

Another reason for expecting that switches are more automatic and effective for established identities is a potential chronic accessibility of highly integrated identities. Amiot et al. (2007) propose that this happens when an individual experiences "repeated exposure to situations that activate the links between the self and one's new group" (Amiot et al., 2007, p. 382). They further explain that this repeated exposure and potential chronic accessibility of identities might lead to a strengthening of connections between novel and existing identities. Thus, switching between two highly integrated, chronically accessible identities might be effective, while switches that include a not-yet-integrated identity might result in identity activation costs.

To our knowledge, no study has directly compared switching away from a minimal group identity (representing a novel identity) to switching away from an established identity towards another established identity. Based on social self-regulation theory (Abrams, 1994) and the model by Amiot et al. (2007), we would expect switches including established identities to be more effective, and (less costly), than switches that include a novel, less integrated identity (e.g., a minimal group identity). This leads to the following hypothesis:

H1. Switching away from a novel identity should result in a smaller congruency effect than switching away from an established identity

A figure with the expected result pattern for the hypothesis is included in Appendix E.

3.1. Methods

3.1.1. Participants and design

We aimed to test a total of $N = 48$[11] participants based on Brysbaert

[10] OSF anonymous pre-registration link for Study 2: https://osf.io/mks85/?view_only=4de5dea90cd04b7ea49674d653ce7c9f

[11] We recruited 48 participants on Prolific Academic. Although 49 participants started the study, one participant stopped early and was therefore reimbursed but their data excluded from the analysis. Another 10 participants showed interest on Prolific but did not start the study - their status showed either "returned" or "timed-out".

and Stevens (2018).[12] In line with the specific IAT tasks used in this study, participants were of British nationality and between 18 and 25 years old ($M = 21.51$, $SD = 2.20$) corresponding to the social identities "young" and "British". The majority of the participants (90%) reported English as their first language, 62.5% participants were female and 37.5% male. All participants had normal or corrected vision; 85% of participants were right-handed and 15% left-handed.

The study has a 2 (congruency: congruent vs. incongruent trials) x 2 (previous identity: established vs. novel) within-subjects design with RTs and error rates as the dependent variables. Each participant completed a series of IATs that included a switch from a minimal group IAT to an age IAT - corresponding to a switch away from a novel identity - and a switch from a nationality IAT to an age IAT - corresponding to a switch away from an established identity. The order of the two switches was balanced over participants (see Fig. 3). Ethical approval was obtained from the [blinded for review] ethics committee.

The main differences to Study 1 are as follows. First, Study 2 does not compare repeat and switch trials. Instead, two different types of switches are compared - a switch away from an established identity vs. a switch away from a novel identity. Second, RTs are compared for a switch towards the same identity (here: age identity), rather than RTs from two different IATs (Study 1: age and race). Although participants will complete a minimal group IAT and a nationality IAT, we are only interested in the RTs of the age identity IATs - the only switch IATs. Third, Study 2 uses Identification IATs instead of the standard IAT as in Study 1. Fourth, Study 2 was administered online.[13]

3.1.2. Materials and measures

Apart from the changes described above, the materials and measures were closely modelled on the previous study, including the use of five successive IATs.

3.1.2.1. Social identities and stimuli. In addition to two established social identities - nationality (British vs. German) and age (young vs. old) - this study included one new identity based on the membership of a "minimal group" (in-group: blue team; out-group: red team).

The study included two sets of face stimuli - one for the nationality and age IAT and one for the minimal group and age IAT. Each stimulus set consisted of 12 faces (6 old faces and 6 young faces - 3 of each age subset were males and 3 were females).[14] For the minimal group and nationality IAT, 6 faces were allocated to the in-group (Blue team or British) and 6 faces to the out-group (Red team or German). The number of old and young faces was the same in the in-group and out-group (3 each). The same stimuli were used in the age IAT, but participants were now asked to categorise them as young faces vs. old. We selected face stimuli from the stimuli used in Study 1 (Bainbridge et al., 2013; Minear & Park, 2004). Based on RTs and accuracies in the age IATs in Pilot 1, we ensured that the two sets of stimuli used in Study 2 had similar mean RTs (Set 1: $M = 799$, $SD = 33.38$; Set 2: $M = 803$, $SD = 34.22$) and accuracies (Set 1: $M = 94.2\%$, $SD = 0.02$; Set 2: $M = 94.0\%$, $SD = 0.02$).

Prior to the nationality IAT and the minimal group IAT, participants were told their group allocation. Since memorization is an effective minimal group induction procedure, especially for implicit measures (Pinter & Greenwald, 2010), we gave participants two minutes to learn the faces of their in-group members (for minimal group IAT: members of the blue team; for nationality IAT: British people). Nationality and minimal group memberships could not be determined visually prior to completing the practice section on the group allocations. After the learning phase, participants were presented with in-group and out-group faces and were asked to select their in-group members. Once all faces were identified correctly, participants were asked to categorise faces appearing on the screen as either in-group or out-group and received feedback on their accuracy. After this practice, participants completed the corresponding nationality or minimal group IAT followed by an age IAT (see Fig. 2).

3.1.2.2. Congruency. Each IAT consisted of 48 congruent trials and 48 incongruent trials; thus, the IATs were slightly shorter than in the previous study to prevent high dropout rates in the online paradigm. The combinations of the following aspects of the design were balanced over participants: whether IATs started with congruent or incongruent trials, whether self-relevant words were allocated (for the entire study) to the left or right finger, which stimulus set was allocated to which IAT, and whether participants started with the minimal group or nationality IAT. Since identification IATs (e.g., Pinter & Greenwald, 2010) rather than standard IATs are used in this study, attributes were self-relevant words (e.g., "I", "my") and other-relevant words (e.g., "they", "others"), which participants are asked to sort into "me" and "not me". On congruent trials, self-relevant words were paired with the in-group whereas on incongruent trials other-relevant related words were paired with the respective in-group.

The IATs were administered online using the Qualtrics software. To create IATs in Qualtrics, we used the web application of the online tool iatgen (Carpenter et al., 2019) and altered the provided code to create successive IATs with the same adaptations of the "standard" IAT as in Study 1.

3.1.2.3. Switch. Participants completed one switch away from an established identity (nationality: British) to another established identity (age: young) - and one switch away from a novel identity (minimal group: blue) to an established identity (age: young) in counterbalanced order. Since we are only interested in comparing the RTs/error rates for switch IATs, we kept the second identity constant (here: age).

3.1.2.4. Demographics and identification questionnaire. [15] The questionnaire included demographic items as well as the 4-item (Doosje et al., 1995) and 1-item (based on Haslam et al., 1999) measures of the strength of identification for national and age identity. The questionnaire also included items assessing interactions with the in- and out-group, and the compatibility of the different groups (whether participants perceived any conflict between the group memberships; based on Benet-Martínez and Haritatos (2005). Those items were included to find out more about any potential design issues and they are not included in the analysis presented here. Finally, participants were asked to indicate on 6-point Likert scales whether they tried to influence their results (1 = not at all, 6 = very much), and how easy they found it to activate the three different social identities (after recoding: 0 = very easy, 5 = very difficult).

3.2. Procedure

Participants were recruited and participant payment was organised through the online platform Prolific Academic. The entire study was run on Qualtrics and took approximately 20–30 min to complete.

Participants were informed that they could only complete the study on a computer and received an error message if they tried to start the Qualtrics survey on a phone. After providing written informed consent participants completed a practice IAT. To avoid already making a social identity salient, the practice IAT asked participants to sort self- vs. other-

[12] For sample size calculations please see supplementary materials S1.2

[13] We disclose all measures, manipulations, and exclusions in this study. We report how we determined the final sample size and whether data collection was continued after data analysis.

[14] In Studies 2 and 3 we only included faces of one racial category (White) to avoid interference from race categorisation

[15] For the full questionnaire and a summary of descriptive statistics not included in this report please see supplementary materials S3

related words and positive vs. negative words (rather than face stimuli as in the main IATs). Prior to each set of main IATs, participants practised the recall of the relevant in-group faces (minimal group or nationality group). Participants then completed the main IATs as described earlier (see also Fig. 4). At the end of the study, participants completed the demographics and identification questionnaire and received debrief information and the financial reimbursement of £4.

3.3. Results

All RTs under 200 ms, practice trials and the practice IAT were excluded from the analysis. Since stimuli did not time out in the online IAT version, all RTs above 3000 ms were excluded. Participants that had 10 or more RTs over 3000 ms in the 48 congruent or 48 incongruent trials of any IAT (over 20% or more very long RTs) were also excluded as this indicated that the participant might have lost focus or might have been distracted from the task. The RT analysis only included correct trials. Two participants were excluded as outliers based on their overall error rates (3 SDs above the mean) and one participant was excluded based on 11 very long RTs (above 3000 ms) in the incongruent trials of one IAT they completed. One participant was excluded because the participant's age was above the limit of 25. To ensure that all conditions were still balanced (see Method: Congruency) we tested 4 additional participants[16] in the same balancing cells as the excluded participants, which ensured the balancing was complete over 48 participants (as planned) in the final analysis.[17]

3.3.1. Minimum precondition for hypothesis testing

As in Study 1, we tested for the minimum precondition that we could detect a significant congruency effect in the IATs that participants completed prior to the switch - specifically, whether RTs were faster and error rates lower for congruent compared with incongruent trials in the minimal group IAT and in the nationality IAT.

A 2 (congruency) x 2 (type of identity) ANOVA for the minimal group and nationality IAT showed the expected main effect of congruency for both RTs ($F(1,47) = 146.43$, $p < .001$, $\eta_p^2 = 0.757$) and error rates ($F(1,47) = 38.25$, $p < .001$, $\eta_p^2 = 0.449$). Participants responded more slowly and made more errors in incongruent compared to congruent trials in both the minimal group IAT and the nationality IAT (see Fig. 5). There was no significant effect of type of identity (RTs: $F(1,47) = 0.11$, $p = .740$, $\eta_p^2 = 0.002$; error rates: $F(1,47) = 1.48$, $p = .230$, $\eta_p^2 = 0.031$).

The interaction between type of identity and congruency was non-significant for RTs ($F(1,47) = 1.46$, $p = .232$, $\eta_p^2 = .030$) but significant for error rates ($F(1,47) = 10.41$, $p = .002$, $\eta_p^2 = .181$). The minimal group IAT showed a lower congruency effect in error rates ($M_{con} = 5.9\%$; $M_{inc} = 9.8\%$; $M_{Diff} = 3.8\%$, $SE = 0.91$) than the nationality IAT ($M_{con} = 5.0\%$; $M_{inc} = 12.5\%$; $M_{Diff} = 7.5\%$, $SE = 1.22$). However, separate follow-up analyses for error rates show that the congruency effect was significant for both type of identities (minimal group: $F(1,47) = 17.82$, $p < .001$, $\eta_p^2 = .275$; nationality: $F(1,47) = 37.77$, $p < .001$, $\eta_p^2 = .446$). The results fulfil the preconditions for hypothesis testing - both identities (minimal group and nationality) showed the expected congruency effect, hence it can be assumed that each identity was made salient during the corresponding IAT.

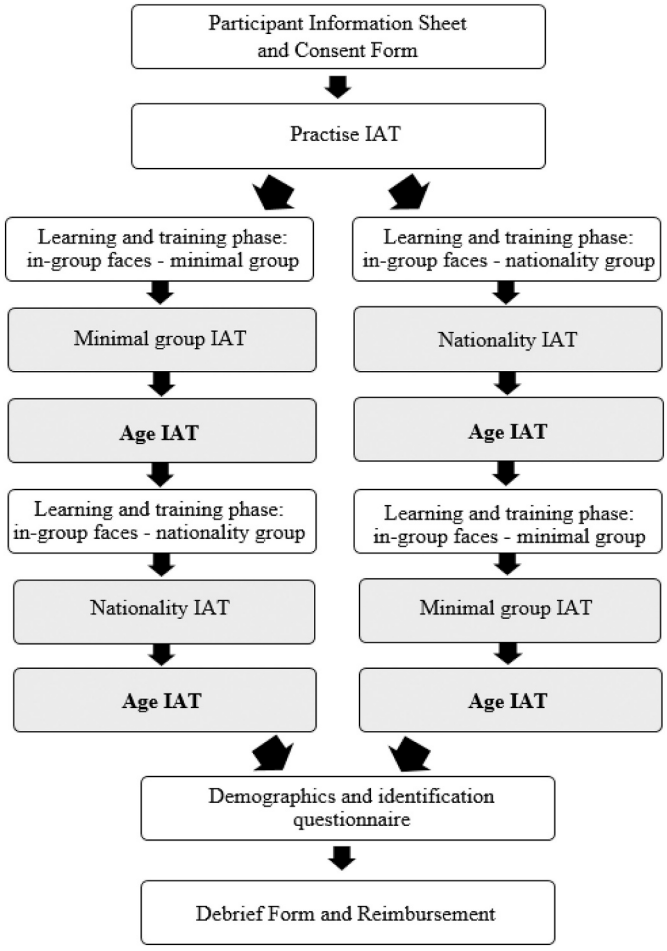

Fig. 4. Study Procedure.
Note. * For the main hypothesis we analysed the age IATs (two for each participant - one after switching away from an established identity and one after switching away from a novel identity).

As a further indicator, we compared the self-reported difficulty to activate the minimal group and national identity. Although the congruency effect shows that both identities were made salient, we found that participants reported significantly greater difficulty ($t(47) = 4.95$, $p < .001$, $d = 0.72$) activating the minimal group identity ($M = 1.73$, $SD = 1.53$) compared to the national identity ($M = 0.73$, $SD = 0.89$). This finding is in line with the novel minimal group identity being less established than the national identity, and it further supports our rationale for comparing the switches away from the two types of identity.

3.3.2. Hypothesis testing

To test whether switches away from a novel identity show a smaller congruency effect than switches away from an established identity, we conducted a 2 (congruency) x 2 (previous identity) ANOVA comparing RTs and error rates in the age IATs where the switch occurred.

We did not find a significant interaction between congruency and previous identity for RTs ($F(1,47) = 2.61$, $p = .113$, $\eta_p^2 = .053$) or error rates ($F(1,47) = 1.25$, $p = .270$, $\eta_p^2 = .026$). For RTs, there was a numerical trend towards the expected smaller congruency effect when switching from the novel identity ($M_{Diff} = 113$, $SE = 18.76$) as compared to switching from the established identity ($M_{Diff} = 157$, $SE = 27.52$). This ~30% difference in the congruency effect seems to be driven by the RT difference on incongruent trials. Error rates showed the reversed (non-significant) pattern of results (Switch from novel identity: $M_{Diff} = 4.46\%$, $SE = 1.12$; Switch from established identity: $M_{Diff} = 2.90\%$, $SE =$

[16] The additional 4 participants were tested after a first analysis of the data as it only became evident at this point that the different conditions were no longer balanced between participants. However, the results did not change substantially. In the registered report we will only stop recruitment once the exact required number of participants has been reached (to ensure all conditions are still balanced).

[17] We recruited the 4 additional participants on Prolific Academic. Due to an issue with the age inclusion criteria 8 participants completed the study of which 3 had to be excluded prior to analysis due to being above 25 years of age and 1 due to stopping the study early.

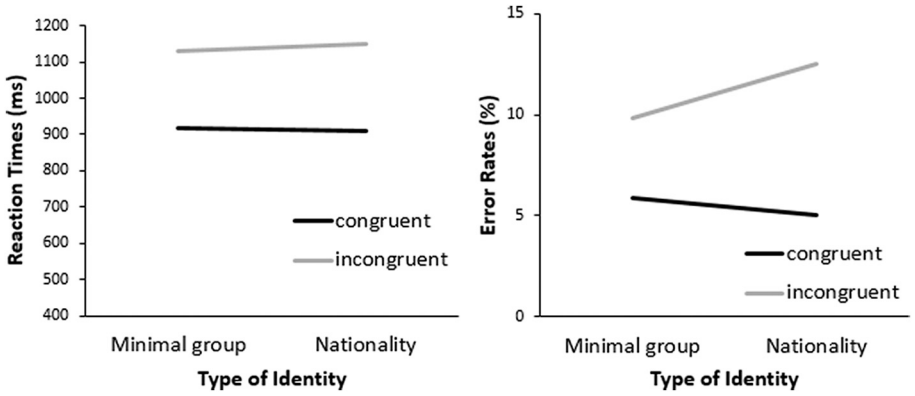

Fig. 5. RTs and error rates as a function of congruency and type of identity.

1.02). We therefore found ambiguous trends concerning Hypothesis 1 (see Fig. 6).

The analysis also revealed the expected main effect of congruency in the age IATs (confirming the activation of the age identity) for both RTs ($F(1,47) = 49.98$, $p < .001$, $\eta_p^2 = 0.515$) and error rates ($F(1,47) = 20.48$, $p < .001$, $\eta_p^2 = 0.303$) with slower responses and higher error rates in incongruent (RTs: $M = 1006$; error rates: $M = 8.3\%$) compared to congruent (RTs: $M = 871$; error rates: $M = 4.6\%$) trials. We found no main effect of previous identity for RTs ($F(1,47) = 3.31$, $p = .075$, $\eta_p^2 = 0.066$) or error rates ($F(1,47) = 0.43$, $p = .515$, $\eta_p^2 = 0.009$).

3.4. Discussion

Study 1 suggested that social identity switches are relatively effective and rapid. Study 2 further investigated social identity switching by comparing switches away from an established identity to switches away from a novel identity. Based on the model by Amiot et al. (2007) and social self-regulation theory (Abrams, 1994), we expected that switches away from a novel identity that has not yet been integrated in the self-concept (minimal group identity) would be less effective than switches between established, more integrated identities.

There was a trend towards a reduced congruency effect when switching away from a novel identity for RTs, but error rates showed the reversed trend; neither trend reached statistical significance. The following, registered, study investigates this ambiguous pattern of results with sufficient power to identify which trend - if any - is supported by further, more ample, evidence. The results from the registered study can help us understand whether the type of social identity impacts on the effectiveness of identity activation during a switch.

A further important aspect, that is addressed in the registered study, is that participants might not have switched identities but rather kept both identities salient concurrently (cross-categorisation theory; Brown & Turner, 1979; Deschamps & Doise, 1978, see above). We tested for this in Study 1 and found little evidence for cross-categorisation for established identities – the registered study allows us to further examine the possibility for cross-categorisation when a novel identity is salient.

4. Study 3 (registered study)

The main aim of the registered study is to further compare the effectiveness of switches away from established vs. from novel identities. A non-significant trend in data in the Study 2 suggests that such a difference cannot yet be ruled out. As already explained in the introduction to Study 2, the findings will test predictions based on social self-regulation theory (Abrams, 1994) and the model by Amiot et al. (2007). We interpret these accounts to suggest that a switch away from a novel social identity may impair activation of an established social identity more than a switch away from another established social identity that is more likely to have been integrated in the self-concept.

The registered study uses the same within-subject design, methods and procedures as Study 2 but with greater statistical power to detect a potentially small effect. The only difference to the materials used in Study 2 is the addition of a question assessing the valence of the different identities, because this is a factor that could drive differences in switching effectiveness. At this point - while we are still determining whether differences in switching effectiveness for novel vs. established identities exist - valence was only included as a descriptive measure, or in exploratory analyses if we detect differences in the effectiveness of switching.

Based on the theory described in Study 2, and the trend in RTs (which was stronger than the opposite trend in the error rates), our main hypothesis remained the same as in Study 2:

H1. Switching away from a novel identity should result in a smaller congruency effect than switching away from an established identity.

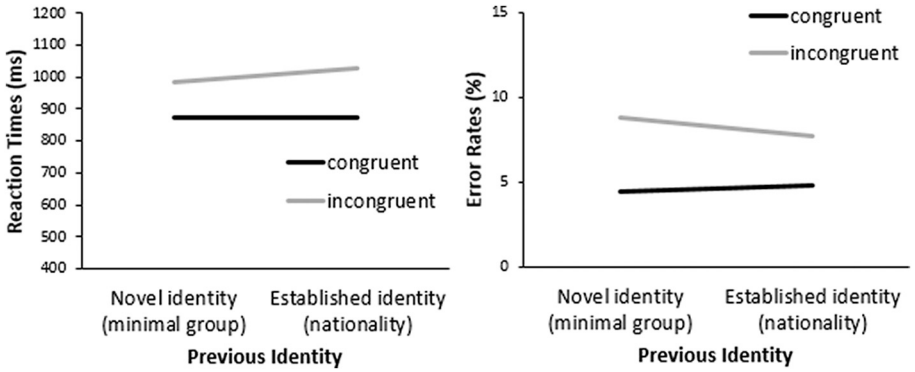

Fig. 6. RTs and error rates as a function of congruency and previous identity.

In addition, the increased statistical power also allows us to test for cross-categorisation (Brown & Turner, 1979; Deschamps & Doise, 1978) as in Study 1. This tests the possibility that participants keep both identities (minimal group and age; nationality and age) salient rather than switching between them.

H2. During a switch, the congruency effect is expected to be reduced more strongly for trials with mixed membership group stimuli as compared to double in- or double-outgroup stimuli.

Appendix F includes graphs with the expected result patterns. In the same way as the second study, the registered study was an online study run on Qualtrics. The study received approval from the departmental ethics committee at the University of [blinded for review].[18]

4.1. Participants

We aimed to recruit at least 144 participants (based on Brysbaert, 2019; Brysbaert & Stevens, 2018)[19] on Prolific Academic and our departmental recruitment system. To take part in the study, participants had to be young (between 18 and 25 years old), British, and be proficient English speakers.

The final sample of 144 participants (44.9% female, 43.7% male, 2.5% other or "prefer not to say", 8.9% missing) was recruited entirely through Prolific Academic. All participants indicated on Prolific that they were proficient English speakers (8.2% were bi- or multi-lingual). On average participants were 22 years old (SD = 2.27) and all included participants indicated being British as one of their nationalities. Further, 78.5% of participants were right-handed and 12.7% left-handed (8.9% missing responses).

4.1.1. Analysis plan

The study was a 2 (congruency) x 2 (previous identity: established vs. novel) within-subjects design. Dependent variables were RTs and error rates. The criteria for excluding data, listed below, are identical to the second study:

- Participants who are not British
- Participants whose age is not between 18 and 25
- Trials with RTs under 200 ms and above 3000 ms
- Participants with 7 or more RTs over 3000 ms in the 48 congruent or 48 incongruent trials of any IAT (15% or more very long RTs) - this criterion is stricter than in Study 1 (over 20%) to ensure that participants that might get distracted from the task throughout the experiment are excluded
- In the RT analysis: trials containing errors
- Participants who are 3SDs above or below the sample mean for RTs or error rates

As planned, for excluded participants, data collection was continued until we reached the required sample size before hypothesis testing commenced.

4.1.2. Registered analyses

4.1.2.1. Testing preconditions. We test for the precondition of a significant congruency effect in both the minimal group IAT and nationality IAT. To test for this, we conduct a 2 (congruency) x 2 (type of identity: minimal group vs. nationality) ANOVA for RTs and error rates to investigate the main effect of congruency and a potential interaction between congruency and type of identity. If we find this interaction, we also compare congruent and incongruent trials for the minimal group IAT and the nationality IAT separately.

4.1.2.2. Testing Hypothesis 1 (identity switching). We conduct a 2 (congruency) x 2 (previous identity: established vs. novel) ANOVA on RTs and error rates in the age IATs to test the main hypothesis of identity switches being more effective when switching away from established as compared to novel identities.

4.1.2.3. Testing Hypothesis 2 (cross-categorisation). As in Study 1, we test for the possibility that - in line with cross-categorisation (Brown & Turner, 1979) - two identities were salient at the same time resulting in participants not having to switch identities. This would be reflected in a significant interaction between congruency and the type of stimuli. We test this separately for the two types of identities (established vs novel) by conducting two 2 (congruency) x 2 (stimuli: mixed membership group vs. double in-or out-group stimuli) ANOVAs for RTs and error rates in the age IATs.

4.2. Results

Based on the inclusion criteria outlined in the analysis plan, we continued collecting data until we reached the required sample size (N = 144)[20] before hypothesis testing commenced. As in the previous studies, we only included correct trials in the RT analyses.

4.2.1. Minimum precondition for hypothesis testing

The 2 (congruency) x 2 (type of identity: minimal group vs. nationality) ANOVAs showed a significant congruency effect for RTs ($F(1,143) = 460.36$, $p < .001$, $\eta_p^2 = 0.763$) and error rates ($F(1,143) = 143.85$, $p < .001$, $\eta_p^2 = 0.501$) in the expected direction, as shown in Fig. 7. As in Study 2, the effect of type of identity was non-significant (RTs: $F(1,143) = 1.21$, $p = .274$, $\eta_p^2 = 0.008$; error rates: $F(1,143) = 0.77$, $p = .383$, $\eta_p^2 = 0.005$).

There was no significant interaction between type of identity and congruency for error rates ($F(1,143) = 0.06$, $p = .809$, $\eta_p^2 = 0.000$). However, the interaction was significant for RTs ($F(1,143) = 4.32$, $p = .039$, $\eta_p^2 = 0.029$) with a stronger RT based congruency effect in the nationality IAT ($M_{con} = 867$; $M_{inc} = 1100$; $M_{Diff} = 232.79$, $SE = 13.42$) compared to the minimal group IAT ($M_{con} = 871$; $M_{inc} = 1074$; $M_{Diff} = 203.65$, $SE = 11.18$). Based on this result, we ran a separate follow-up analysis that revealed a significant RT-based congruency effect for both the minimal group identity ($F(1,143) = 331.93$, $p < .001$, $\eta_p^2 = 0.699$) and national identity ($F(1,143) = 300.68$, $p < .001$, $\eta_p^2 = 0.678$). As in Study 2, those findings support that both identities were salient in the experiment and therefore fulfil the preconditions for hypothesis testing.

Hypothesis 1. Identity switching.

The 2 (congruency) x 2 (previous identity: established vs. novel) ANOVA in the age IATs showed no significant interaction between congruency and previous identity (RTs: $F(1,143) = 1.44$, $p = .232$, $\eta_p^2 = 0.010$; error rates: $F(1,143) = 0.03$, $p = .861$, $\eta_p^2 = 0.000$) and therefore no support for Hypothesis 1 - identity switches being more effective when switching away from established as compared to novel identities. As shown, in Fig. 8, we found the expected main effect of congruency (RTs: $F(1,143) = 131.66$, $p < .001$, $\eta_p^2 = 0.479$; error rates: $F(1,143) = 102.06$, $p < .001$, $\eta_p^2 = 0.416$) which supports that participants did

[18] We disclose all measures, manipulations, and exclusions in this study. We report how we determined the final sample size and whether data collection was continued after data analysis.

[19] For sample size calculations please see supplementary materials S1.3

[20] We continued data collection until we reached 144 participants that meet all inclusion criteria as listed in the results section. To achieve this, we had to exclude and retest 19 participants (3 did not meet the study inclusion criteria; 2 due to issues with data recording or balancing; 8 for being 3*SD*s above or below the sample mean for RTs or error rates; 6 due to prolonged RTs in one or more blocks). The data was only analysed once data collection had been completed.

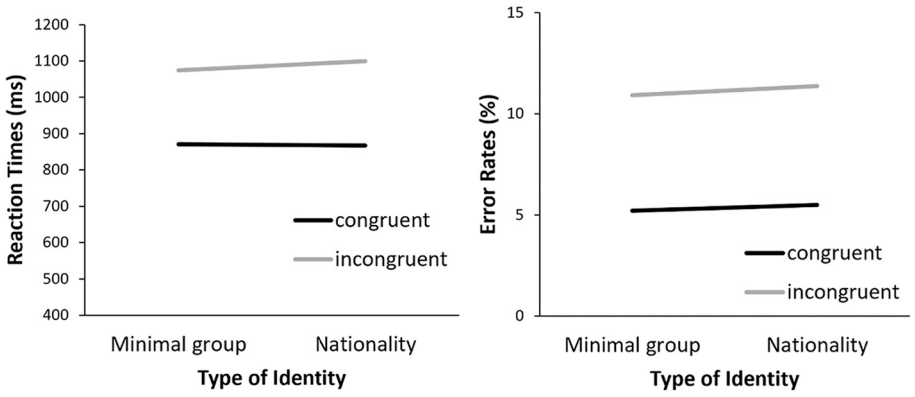

Fig. 7. Study 3: RTs and error rates as a function of congruency and type of identity.

activate their age identity. The main effect of the previous identity was non-significant for both RTs ($F(1,143) = 1.75$, $p = .188$, $\eta_p^2 = 0.012$) and error rates ($F(1,143) = 0.26$, $p = .611$, $\eta_p^2 = 0.002$).

Hypothesis 2. Cross-categorisation.

An three-way ANOVA showed a significant congruency x stimulus type (mixed membership group vs. double in-or out-group membership) x previous identity (minimal group vs. nationality) interaction for RTs ($F(1,143) = 5.64$, $p = .019$, $\eta_p^2 = 0.038$) and a marginally significant interaction for error rates ($F(1,143) = 2.76$, $p = .099$, $\eta_p^2 = 0.019$). These results justify our planned analysis to test for cross-categorisation separately for the two types of identities (established vs novel). For the age IAT following the minimal group IAT, we found a significant interaction between congruency and stimulus type for RTs ($F(1,143) = 6.09$, $p = .015$, $\eta_p^2 = 0.041$) and error rates ($F(1,143) = 7.17$, $p = .008$, $\eta_p^2 = 0.048$). As shown in Fig. 9, the findings support Hypothesis 2 (testing for cross-categorisation) with a larger congruency effect for double in- or out-group membership (RTs: $M_{con} = 799$; $M_{inc} = 916$; $M_{Diff} = 117.28$, $SE = 15.45$; errors: $M_{con} = 4.1\%$; $M_{inc} = 11.2\%$; $M_{Diff} = 7.1\%$, $SE = 1.10$) as compared to mixed membership (RTs: $M_{con} = 816$; $M_{inc} = 899$; $M_{Diff} = 83.25$, $SE = 13.51$; errors: $M_{con} = 6.0\%$; $M_{inc} = 9.4\%$; $M_{Diff} = 3.5\%$, $SE = 1.02$). We also found the expected main effect of congruency (RTs: $F(1,143) = 61.69$, $p < .001$, $\eta_p^2 = 0.301$; error rates: $F(1,143) = 42.47$, $p < .001$, $\eta_p^2 = 0.229$) and no main effect of stimulus type (RTs: $F(1,143) = 0.001$, $p = .973$, $\eta_p^2 < 0.001$; error rates: $F(1,143) = 0.02$, $p = .902$, $\eta_p^2 < 0.001$) in the age IAT following the minimal group IAT. For the age IAT following the nationality IAT, we also found a significant interaction between congruency and stimuli for error rates ($F(1,143) = 23.91$, $p < .001$, $\eta_p^2 = 0.143$) in the expected direction as shown in Fig. 10 and therefore supporting Hypothesis 2 (mixed membership: $M_{con} = 6.4\%$; $M_{inc} = 8.7\%$; $M_{Diff} = 2.2\%$, $SE = 1.09$; double in- or out-group membership: $M_{con} = 3.0\%$; $M_{inc} = 12.3\%$; $M_{Diff} = 9.23\%$, $SE = 1.07$). However, for RTs, the interaction was non-significant ($F(1,143) = 0.69$, $p = .409$, $\eta_p^2 = 0.005$). Further, we again found a main effect of congruency (RTs: $F(1,143) = 75.45$, $p < .001$, $\eta_p^2 = 0.345$; error rates: $F(1,143) = 50.47$, $p < .001$, $\eta_p^2 = 0.261$) and no main effect of stimulus type (RTs: $F(1,143) = 2.58$, $p = .110$, $\eta_p^2 = 0.018$; error rates: $F(1,143) = 0.02$, $p = .894$, $\eta_p^2 < 0.001$) in the age IAT following the nationality IAT.

4.2.2. Exploratory analyses

4.2.2.1. Self-reported difficulty to activate identity and identity valence. [21] Next to the implicit measures assessed in our main hypotheses, we explicitly asked participants how difficulty they found it to activate the minimal group and national identity. Participant's self-report showed that they found it significantly more difficult ($t(141) = 6.91$, $p < .001$, $d = 0.58$) to activate the minimal group identity ($M = 2.18$, $SE = 0.11$) compared to the national identity ($M = 1.31$, $SE = 0.10$). Further, we asked participants to indicate how positive they felt about the three different identities (valence of the identities; Cruwys et al., 2016). Participants rated the national identity ($M = 5.99$, $SE = 0.22$) as significantly more positive ($t(136) = -4.02$, $p < .001$, $d = 0.34$) to themselves as the minimal group identity ($M = 5.06$, $SE = 0.18$). These findings indicate that the novel identity has not been established within the self yet, showing that we had good reasons for comparing switches between the two types of identities (see Study 2).

4.3. Discussion

In the registered study we replicated Study 2 with an increased sample size to establish whether there are differences in effectiveness of switching away from established vs. novel identities. Based on social self-regulation theory (Abrams, 1994) and the four-stage model of social identity development and integration (Amiot et al., 2007), we expected that switching from a novel social identity to an established identity would be less effective than switching from one established identity to another. We found no support for a reduced congruency effect when switching away from a novel identity compared to an established identity for RTs and error rates. The results suggest that the level of integration in the self of a social identity does not affect the effectiveness of identity switching.

Further, the increased sample size in this study allowed us to test for cross-categorisation (Brown & Turner, 1979; Deschamps & Doise, 1978) – both identities remaining salient concurrently. We expected that – if both identities were salient – the congruency effect would be stronger for face stimuli that hold either double in- or out-group attributes as compared to stimuli that hold some in- and some out-group attributes (mixed membership stimuli). We found the expected pattern for cross-categorisation for switches away from a novel identity for both RTs and error rates. However, when switching away from an established identity, we only found support for cross-categorisation in the pattern of

[21] For a summary of descriptive statistics not included in this report, as well as changes to the Demographics and Identification Questionnaire compared to study 2, please see supplementary materials S4

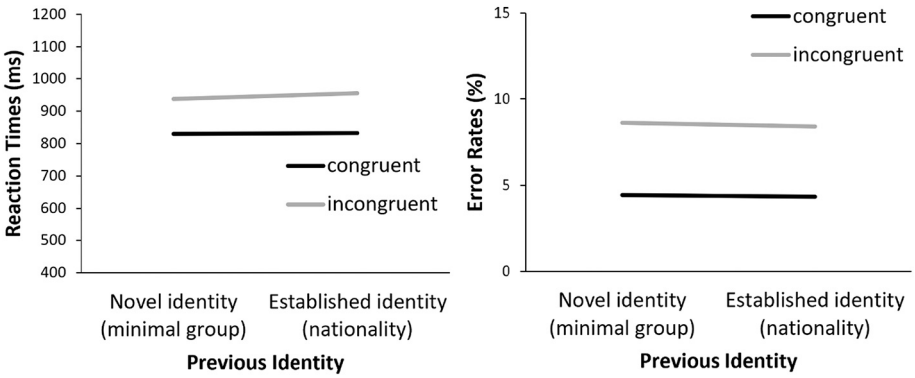

Fig. 8. Study 3: RTs and error rates as a function of congruency and previous identity.

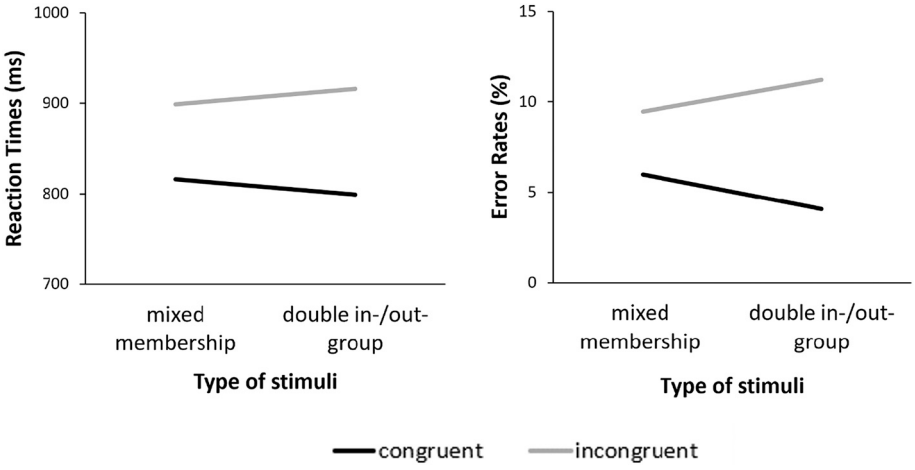

Fig. 9. Switch from novel identity (minimal group): Age IAT RTs and error rates as a function of congruency and stimulus type.

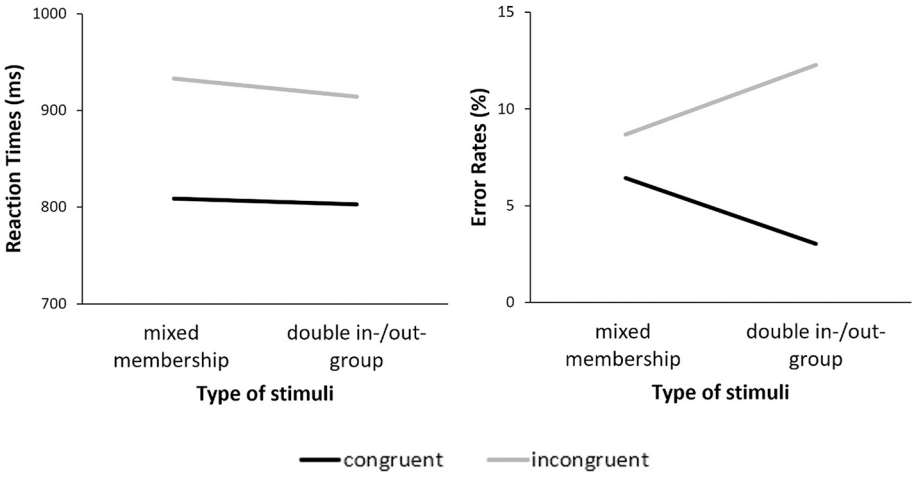

Fig. 10. Switch from established identity (nationality): Age IAT RTs and error rates as a function of congruency and stimulus type.

error rates, but not RTs.

In the following section, we discuss the combined findings of the three presented studies on the effectiveness of social identity switches, switches between different types of identities, and cross-categorisation theory. We also discuss theoretical and practical implications as well as limitations of the presented studies.

5. General discussion

The main aim of the current investigation was to determine the effectiveness of social identity switching. We assessed whether social identity switches lead to identity activation costs (impaired activation of the next identity) and whether social identity switches are less effective for novel than for well-established identities. The absence of an identity activation costs in our results indicates that identity switching is

effective. This has important theoretical implications by lending empirical support to self-categorisation theory that states that social identity switches are "inherently variable, fluid, and context dependent" (Turner et al., 1994, p. 454).

To our knowledge, our investigation is the first approach that has employed key aspects of the task switching paradigm to learn about the process of social identity switching. The potential cost of an identity switch also has important practical implications. Like task switches, social identity switches are ubiquitous. Technological developments over the last decades have resulted in different social identities being only "a click away" from becoming salient. We can interact with (and receive) information about different social identities on a permanent basis, wherever we are - by scrolling through social media, reading news on our smartphone, receiving emails and instant messages, often in rapid succession. The literature on task switch costs has changed the way we view "multi-tasking" by providing a better understanding of its impact on task performance and task selection. Similarly, our research has important practical implications for how well people can deal with frequent and rapid social identity switches.

In our first study, we focused on comparing relatively rapid switches between two well-established identities (age and race) to staying in the same identity. We did not find evidence of an identity activation cost – impaired identity activation during identity switches as comparted to identity repeats. This suggests that switching between two identities is highly effective. In our second study and the registered report, we further tested whether the effectiveness of social identity switches differs between well-established and novel identities. This has potential theoretical implications because it tests for possible differences between newly-acquired and well-established identities as outlined by social self-regulation theory (Abrams, 1994) and the identity integration model by Amiot et al. (2007). Taken together, our findings suggest that switching from a novel identity to an established identity may be as effective as switching between two well-established identities. Results from Study 2 and the registered report are consistent with the theoretical view that social identities are overall highly flexible and context-dependent (e.g., Turner et al., 1994). At least for the identities under our scrutiny, we found little/no support that self-categorisation differs for novel vs. well-established identities (as suggested by e.g., Abrams, 1994; Amiot et al., 2007). Instead, our findings are more in line with previous studies (e.g., Otten & Moskowitz, 2000; Otten & Wentura, 1999) that found minimal group processes to be highly automated.

Importantly, we also tested for cross-categorisation in Study 1 as well as the registered report to account for participants keeping both identities salient - such concurrent salience potentially reducing the need to switch between identities. This analysis has theoretical implications for the competing accounts of functional antagonism (e.g., Turner & Oakes, 1986) and cross-categorisation theory (Brown & Turner, 1979). For the identities in Study 1, our findings did not support a cross-categorisation hypothesis that during a switch double in- and out-group stimuli (White and young; Black and old faces) were more strongly associated with positive and negative words, respectively, than mixed membership stimuli (young and Black; old and White). Indeed, the numerical trends for RTs and error rates were contradictory, suggestive of speed-accuracy trade-offs. To shed more light on these ambiguous results, we tested for cross-categorisation in the registered report. We found support for a co-activation of both identities when participants switched from the minimal group identity to the age identity (both for RTs and error rates), and partial support (in the error rates) for cross-categorisation when participants switched from their national identity to the age identity. As we explain below, certain features of the design of the registered report may have inadvertently made both social identities salient concurrently.

5.1. Limitations

One important difference between Study 1 (in which we found no consistent support for cross-categorisation) and the registered report (in which we found at least partial support for cross-categorisation) is that the registered report included a relatively extensive training and feedback phase during which participants learned the faces of in-group category members (from the minimal group or nationality categorisation). This learning likely led to the formation of associative bindings between in-group faces and the identity relevant in the first IAT (minimal group identity or the national identity). Thus, the presentation of in-group faces during the age IAT may have resulted in associative retrieval of the previously (but no longer) relevant identity – which likely explains the cross-categorisation effect. Crucially, in Study 1 the categorisation of a face (by age or race) did not require learning the associations between individual faces and groups – hence, there was no detectable cross-categorisation effect there.

Effects of the previous "associative history", such as those described above, have been extensively documented in task switching where it was shown that stimuli can re-activate irrelevant task-sets via previously-formed associative bindings. For example, in a study where participants were required to switch between two tasks, Waszak and colleagues (Waszak et al., 2003, 2004, 2005) have presented some stimuli in only one task and other stimuli in both tasks. Performance was better for stimuli previously encountered only in the currently relevant task compared to stimuli previously encountered in both the currently relevant and the currently irrelevant tasks. This was presumably because the latter stimuli associatively retrieved not only the relevant task-set, but also the irrelevant task-set, resulting in task-set competition. More recently, Graham and Lavric (2021) investigated the effect of associative history by examining whether the most recent encounter with the stimulus was in the context of the currently relevant task or in the context of the currently irrelevant task, and found that performance was worse for the latter (they also found this effect in a language-switching paradigm which employed the same stimuli). Although, the earlier studies by Waszak et al. (2003, 2004, 2005) suggested that the associative history effects were greater for task switches than for task repetitions, Graham and Lavric (2021) did not find that switches were more susceptible; they reported comparable associative history effects for switches and repetitions. Whether associative history effects are larger for switches or not, these (and other) studies agree that associative history has robust overall effects on performance.

These findings from task switching indicate that associative history may well explain the cross-categorisation effects observed in the registered report, especially given their absence in Study 1, where learning of the stimulus-category associations was not necessary for performing the IATs – because in this study the categorisations could be done effectively based on facial features related to age/race. However, further research is needed to confirm this conjecture that cross-categorisation effects are indeed limited to memory-based categorisations were stimuli have to be learned, and are not observed for perceptual categorisations where the learning of stimuli is not essential for task performance and/or encouraged by the experimental instructions.

A further potential limitation of the presented research is that in real life people perform more rapid social identity switches than those we induced, and that their effects may be more short-lived. It is therefore possible that we could have observed identity activation costs had we elicited more rapid switches and had we been able to measure identity activation over an even shorter period of time after the identity switch. Our measure of identity activation is at a time-scale of minutes, whereas in task switching research, individual trials typically last seconds.

Notably, we were able to detect task-switch costs over longer task (IAT) blocks lasting minutes, but not identity activation costs. The relatively low temporal resolution of the measurement in our design (compared to that in task switching) was dictated by the IAT-derived measure of identity activation (based on the difference between congruent and incongruent trials). One very promising, yet challenging, future development would be to develop "faster" (higher temporal resolution) measures of identity activation in order to examine the immediate consequences of identity switches. A first step might be to develop a paradigm using fMRI – which may provide a latent measure of identity activation at a timescale of several seconds rather than minutes. Molenberghs and Louis (2018) review previous fMRI studies that have focused on questions such as how activity in different brain regions is influenced by group memberships and how this is linked to in-group bias. Future research could build on such studies and develop a paradigm that clearly distinguishes different identities based on activation patterns in different brain regions, with the potential to monitor faster changes in salience.

5.2. Future research

In the presented studies, we chose identities that were low in potential conflict (e.g., the minimal group identity and national identity). We also did not include negative or stigmatised identities. It is conceivable that identity activation costs might occur for identities that are highly incompatible (e.g., where an identity strongly differs in norms and values from another social identity a person holds) or when the switch includes a negatively valenced identity. Finally, study 2 and the registered report focused on comparing switching from a novel to an established identity vs. switching between two established identities. Yet, the less automatic component of activating novel identities might only become evident when one has to switch from an established identity to a novel identity. To summarise, while our findings show that an identity being novel does not lower the effectiveness of switches per se, future studies should further investigate compatibility, the valence of identities, and the order of identities (novel to established or established to novel) influence the effectiveness of social identity switching.

A further step for future research will be to investigate whether social identity switches lead to performance costs. Our studies focused on identity activation costs and therefore the question whether identity switches lead to an impairment of activating the next identity. We did not find identity activation costs. However, we did not test directly for performance costs such as a slowdown or an increased propensity to make errors.

Finally, other important questions for future research are whether there are individual differences in the ability to switch between social identities, and whether such differences in switching ability relate to well-being outcomes. Earlier on, we discussed how technological developments can result in people having to perform frequent and rapid switches. For instance, switching frequently between one's home and work identity might come with no costs for individuals that tend to switch identities very effectively. However, those who switch identities less effectively might experience higher levels of stress and lower life satisfaction if they are required to switch frequently between identities. In this sense, "switching off from work" may in fact be "effectively switching between" a work and parent or personal identity. Research by Cruwy's and colleagues (2016) shows that an intervention that increased the number of important and positive social identities as well as the compatibility of identities led to improvements in mental health. Similarly, helping people navigate between their different social identities more effectively might benefit well-being.

In a similar vein, future studies could investigate whether frequent – and effective - social identity switches might promise some benefits for outcome variables such as creativity. As part of their model on dual-identities and creativity, Gocłowska and Crisp (2014) suggests that switching between contrasting identities might foster processes linked to creativity and problem solving. Consistent with this suggestion, a study on "code-switching" in bilinguals found that participants who frequently switched languages in the same speech episode had higher innovative capacity than the less frequent code-switchers (Kharkhurin & Wei, 2015).

6. Conclusion

To conclude, our research suggests that social identity switches are highly effective – in line with self-categorisation theory (e.g., Turner et al., 1994). We also found no evidence that switching from a novel to an established identity is less effective than switching between two well-established identities. Future research is needed to further refine the paradigm used in the second set of studies to specify the conditions where cross-categorisation may occur and examine the effectiveness of social identity switches for other (e.g., negatively-valenced or stigmatised) identities.

Open science practices

Studies 1, 2 and the registered report have been pre-registered on OSF- including hypotheses and planned analyses.

- Study 1: https://osf.io/4v3gu
- Study 2: https://osf.io/mks85
- pre-registered Stage 1 report and analysis plan: https://osf.io/p2y5d
- Registered report: https://osf.io/gd7fe

Data access statement

Our study materials as well as the data are openly available on OSF under the following links:

- Pilot 1 and Study 1: https://osf.io/vg8fc/ [DOI: 10.17605/OSF.IO/VG8FC]
- Study 2: https://osf.io/vecgq/ [DOI: 10.17605/OSF.IO/VECGQ]
- Registered Report: https://osf.io/uw63h/ [DOI 10.17605/OSF.IO/UW63H]

Declarations of interest

The authors declare no conflict of interest.

Funding and role of funding source

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK, grant ref.: 2074877 and EP/S001409/1. EPSRC was not involved in the study design, data collection, analysis or interpretation of the data; in the writing of the report; and in the decision to submit the article for publication.

Acknowledgements

We thank Jolanda Jetten as well as the editor, and reviewers for their helpful comments on an earlier version of the manuscript.

Appendix A. Study 1 – expected result patterns for each hypothesis

H1a. The congruency effect is reduced for switches between different salient identities compared to repeating the same salient identity, indicating higher identity activation costs during a switch.

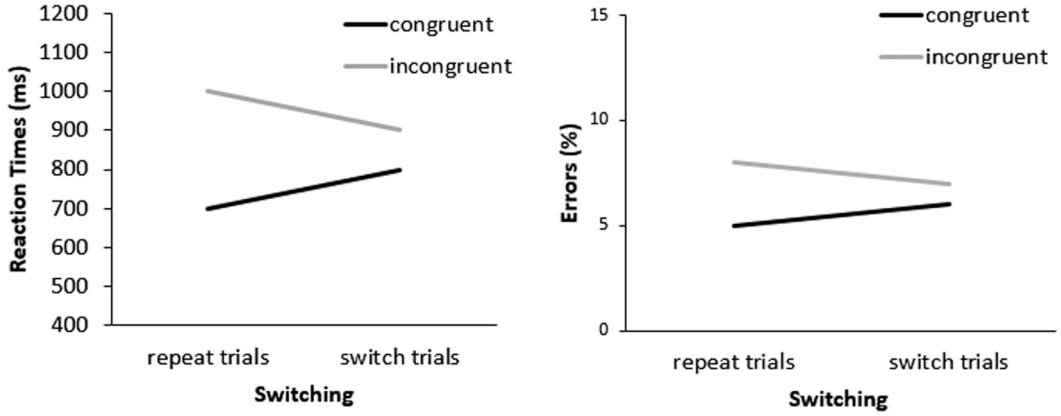

Fig. A1. Expected result patterns for H1a.
Note. We have no strong reason to expect that identity activation costs would be driven more strongly by either congruent or incongruent trials.

H1b. (Null Hypothesis): The congruency effect is equivalent in size for switches between different salient identities compared to repeating the same salient identity, indicating no effect of switching on identity activation.

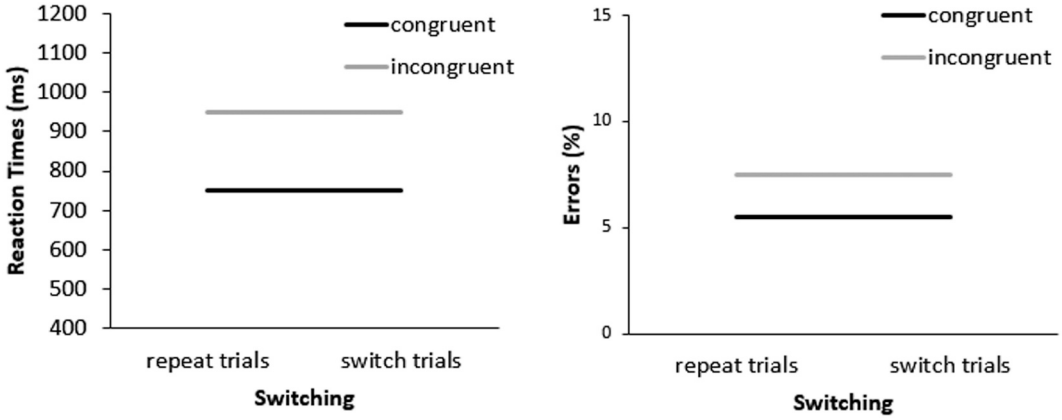

Fig. A2. Expected result patterns for H1b.

H2. Switching between identities will affect the congruency effect more strongly in the first part of the IAT than in the second part of the IAT.

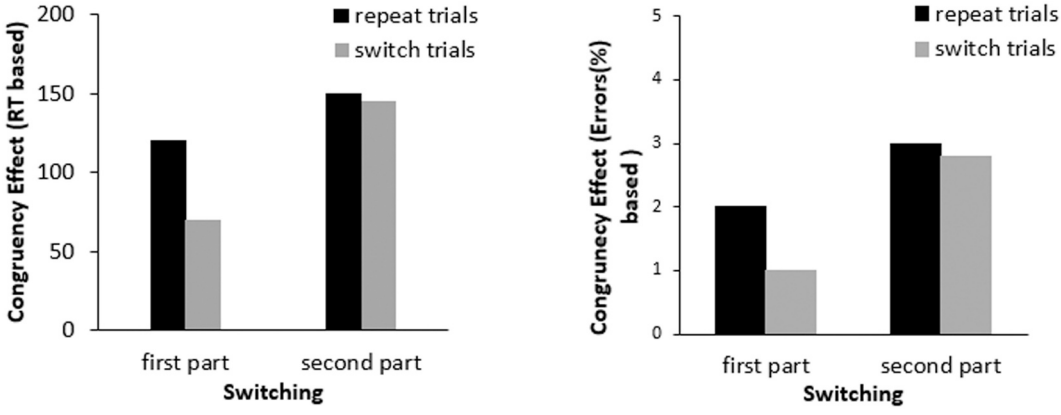

Fig. A3. Expected result patterns for H2.

H3. Switching between salient identities is expected to reduce the congruency effect more strongly for trials with mixed membership group stimuli as compared to double in- or double out-group stimuli.

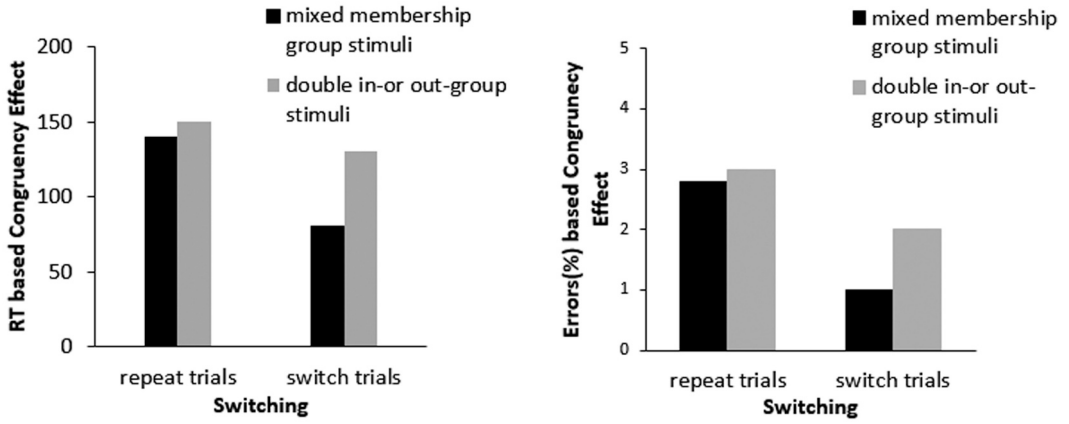

Fig. A4. Expected result patterns for H3.

Appendix B. Pilot 1 - results for the effects of identity, congruency, and switching

A total of 24 participants[22] were included in Pilot 1. There were no outliers in overall RTs and error rates. However, based on spotting outliers using boxplots for congruent and incongruent trials separately in each IAT, 3 sets of RT trials were excluded (2 for participant 18, and 1 for participant 22).

We found the expected congruency effect for both RTs (for results of all significance tests see Table A1) and error rates (see Table A2) as well as a significant interaction between congruency and identity. Participants overall responded faster ($M_{Age} = 820$; $M_{Race} = 855$) and with less errors ($M_{Age} = 3.3\%$; $M_{Race} = 4.5\%$) in congruent compared to incongruent trials (RTs: $M_{Age} = 960$, $M_{Race} = 893$; error rates: $M_{Age} = 7.6\%$, $M_{Race} = 6.5\%$). Importantly, separate follow-up analyses for age and race supported that the congruency effect remained significant for both identities for RTs (Age: $F(1,22) = 117.79$, $p < .001$, $\eta_p^2 = 0.843$; Race: $F(1,23) = 5.02$, $p = .035$, $\eta_p^2 = 0.179$) and error rates (Age: $F(1,23) = 37.92$, $p < .001$, $\eta_p^2 = 0.622$; Race: $F(1,23) = 6.17$, $p = .021$, $\eta_p^2 = 0.212$).

There was a significant effect of switching in RTs (but not error rates) with participants responding faster in repeat ($M = 868$) compared to switch ($M = 895$) trials. For both RTs and error rates there was no significant effect of type of identity and no significant interactions between identity and switching, congruency and switching, and identity, congruency and switching.

Table A1
ANOVA results with RTs as dependent variable.

Effect	*F*	*df*	*p*	η_p^2
Congruency	55.50	1, 22	<0.001	0.716
Switching	5.82	1, 22	0.025	0.209
Identity	1.33	1, 22	0.262	0.057
Congruency x Switching	0.09	1, 22	0.766	0.004
Congruency x Identity	19.95	1, 22	<0.001	0.476
Switching x Identity	1.18	1, 22	0.289	0.051
Congruency x Switching x Identity	0.001	1, 22	0.979	. <0.001

Table A2
ANOVA results with error rates as dependent variable.

Effect	*F*	*df*	*p*	η_p^2
Congruency	26.60	1, 23	<0.001	0.536
Switching	0.002	1, 23	0.962	<0.001
Identity	0.01	1, 23	0.931	<0.001
Congruency x Switching	1.33	1, 23	0.260	0.055
Congruency x Identity	7.51	1, 23	0.012	0.246
Switching x Identity	0.03	1, 23	0.858	0.001
Congruency x Switching x Identity	0.01	1, 23	0.923	. <0.001

[22] We tested until we reached 24 participants. 25 participants completed the study of which one did not meet the study inclusion criteria. The data of this participant was therefore not included in any analysis or data summary. The data was only analysed once data collection had been completed.

Appendix C. Balancing conditions Study 1

Balancing the five factors listed below results in 32 conditions (see Table A3). In E Prime it was also randomised which set of face and word stimuli was allocated to which IAT.

- To which finger "positive" and "negative" were allocated
- The order of race and age IATs
- The order of repeat and switch IATs
- Whether the Startup IAT starts with congruent or incongruent trials
- The order of four main IATs - starting with a congruent or an incongruent response mapping

Order of type of IATs and therefore of switches (sw) and repeats (rep)					Finger allocation	Congruent (Con) or incongruent (Incon) trials first				
Startup IAT	IAT 1	IAT 2	IAT 3	IAT 4	Throughout	Startup IAT	IAT 1	IAT 2	IAT 3	IAT 4
Race IAT	Age (sw)	Age (rep)	Race (sw)	Race (rep)	left positive	Con	Con	Incon	Incon	Con
Race IAT	Age (sw)	Age (rep)	Race (sw)	Race (rep)	left positive	Con	Incon	Con	Con	Incon
Age IAT	Race (sw)	Race (rep)	Age (sw)	Age (rep)	left positive	Con	Incon	Con	Con	Incon
Age IAT	Race (sw)	Race (rep)	Age (sw)	Age (rep)	left positive	Con	Con	Incon	Incon	Con
Age IAT	Age (rep)	Race (sw)	Race (rep)	Age (sw)	left positive	Con	Incon	Incon	Con	Con
Age IAT	Age (rep)	Race (sw)	Race (rep)	Age (sw)	left positive	Con	Con	Con	Incon	Incon
Race IAT	Race (rep)	Age (sw)	Age (rep)	Race (sw)	left positive	Con	Con	Con	Incon	Incon
Race IAT	Race (rep)	Age (sw)	Age (rep)	Race (sw)	left positive	Con	Incon	Incon	Con	Con
Race IAT	Age (sw)	Age (rep)	Race (sw)	Race (rep)	right positive	Con	Con	Incon	Incon	Con
Race IAT	Age (sw)	Age (rep)	Race (sw)	Race (rep)	right positive	Con	Incon	Con	Con	Incon
Age IAT	Race (sw)	Race (rep)	Age (sw)	Age (rep)	right positive	Con	Incon	Con	Con	Incon
Age IAT	Race (sw)	Race (rep)	Age (sw)	Age (rep)	right positive	Con	Con	Incon	Incon	Con
Age IAT	Age (rep)	Race (sw)	Race (rep)	Age (sw)	right positive	Con	Incon	Incon	Con	Con
Age IAT	Age (rep)	Race (sw)	Race (rep)	Age (sw)	right positive	Con	Con	Con	Incon	Incon
Race IAT	Race (rep)	Age (sw)	Age (rep)	Race (sw)	right positive	Con	Con	Con	Incon	Incon
Race IAT	Race (rep)	Age (sw)	Age (rep)	Race (sw)	right positive	Con	Incon	Incon	Con	Con
Race IAT	Age (sw)	Age (rep)	Race (sw)	Race (rep)	left positive	Incon	Con	Incon	Incon	Con
Race IAT	Age (sw)	Age (rep)	Race (sw)	Race (rep)	left positive	Incon	Incon	Con	Con	Incon
Age IAT	Race (sw)	Race (rep)	Age (sw)	Age (rep)	left positive	Incon	Incon	Con	Con	Incon
Age IAT	Race (sw)	Race (rep)	Age (sw)	Age (rep)	left positive	Incon	Con	Incon	Incon	Con
Age IAT	Age (rep)	Race (sw)	Race (rep)	Age (sw)	left positive	Incon	Incon	Incon	Con	Con
Age IAT	Age (rep)	Race (sw)	Race (rep)	Age (sw)	left positive	Incon	Con	Con	Incon	Incon
Race IAT	Race (rep)	Age (sw)	Age (rep)	Race (sw)	left positive	Incon	Con	Con	Incon	Incon
Race IAT	Race (rep)	Age (sw)	Age (rep)	Race (sw)	left positive	Incon	Incon	Incon	Con	Con
Race IAT	Age (sw)	Age (rep)	Race (sw)	Race (rep)	right positive	Incon	Con	Incon	Incon	Con
Race IAT	Age (sw)	Age (rep)	Race (sw)	Race (rep)	right positive	Incon	Incon	Con	Con	Incon
Age IAT	Race (sw)	Race (rep)	Age (sw)	Age (rep)	right positive	Incon	Incon	Con	Con	Incon
Age IAT	Race (sw)	Race (rep)	Age (sw)	Age (rep)	right positive	Incon	Con	Incon	Incon	Con
Age IAT	Age (rep)	Race (sw)	Race (rep)	Age (sw)	right positive	Incon	Incon	Incon	Con	Con
Age IAT	Age (rep)	Race (sw)	Race (rep)	Age (sw)	right positive	Incon	Con	Con	Incon	Incon
Race IAT	Race (rep)	Age (sw)	Age (rep)	Race (sw)	right positive	Incon	Con	Con	Incon	Incon
Race IAT	Race (rep)	Age (sw)	Age (rep)	Race (sw)	right positive	Incon	Incon	Incon	Con	Con

Table A3. Balancing conditions Pilot 1 and Study 1.

Appendix D. Balancing conditions Study 2 and Study 3

The following factors are balanced between participants:

- Whether the IATs start with congruent or incongruent trials
- To which finger "me" and "not me" are allocated
- Which sets of IATs participants start with (minimal group IAT or nationality IAT)
- The order of the two face stimuli sets

Balancing those factors results in 16 conditions (see Table A4).

IAT starts with congruent trials								IAT starts with incongruent trials							
"me" allocated to left				"me" allocated to right				"me" allocated to left				"me" allocated to right			
Minimal Group IAT first		Nationality IAT first		Minimal Group IAT first		Nationality IAT first		Minimal Group IAT first		Nationality IAT first		Minimal Group IAT first		Nationality IAT first	
A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B

Table A4. Balancing conditions Study 2 and Study 3.
Note. "A" and "B" refers to the type of picture set that is used in the first IAT.

Appendix E. Study 2 – expected result patterns

H1. Switching away from a novel identity should result in a smaller congruency effect than switching away from an established identity.

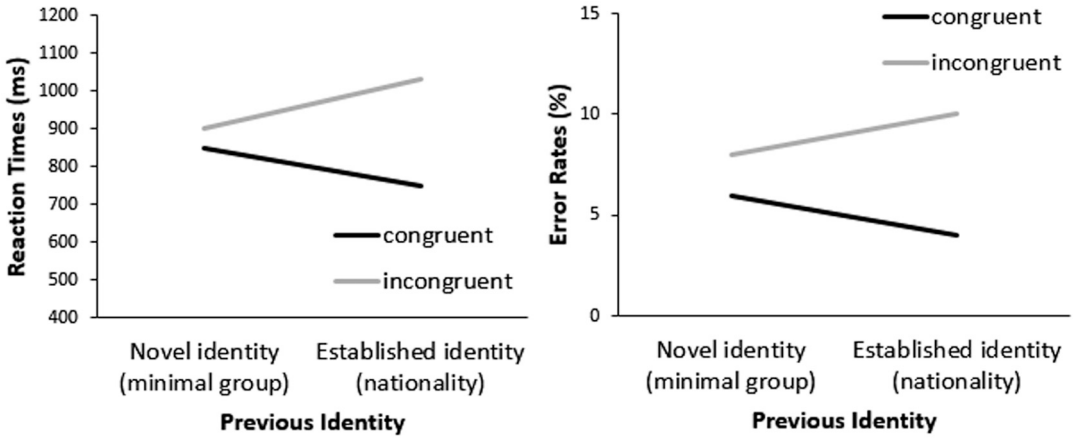

Fig. A7. Expected result patterns for H1

Appendix F. Registered study – expected result patterns

H1. Switching away from a novel identity should result in a smaller congruency effect than switching away from an established identity.

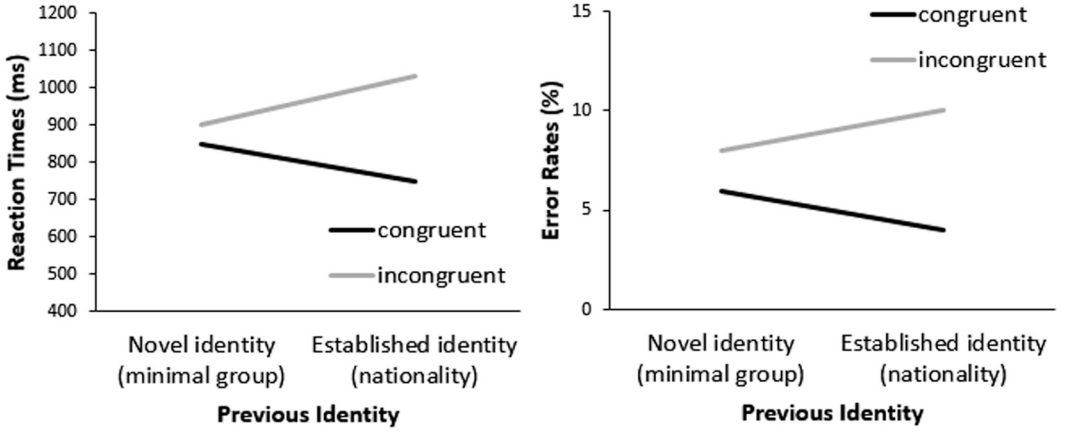

Fig. A8. Expected result patterns for H1.

H2. During a switch, the congruency effect is expected to be reduced more strongly for trials with mixed membership group stimuli as compared to double in- or double-outgroup stimuli.

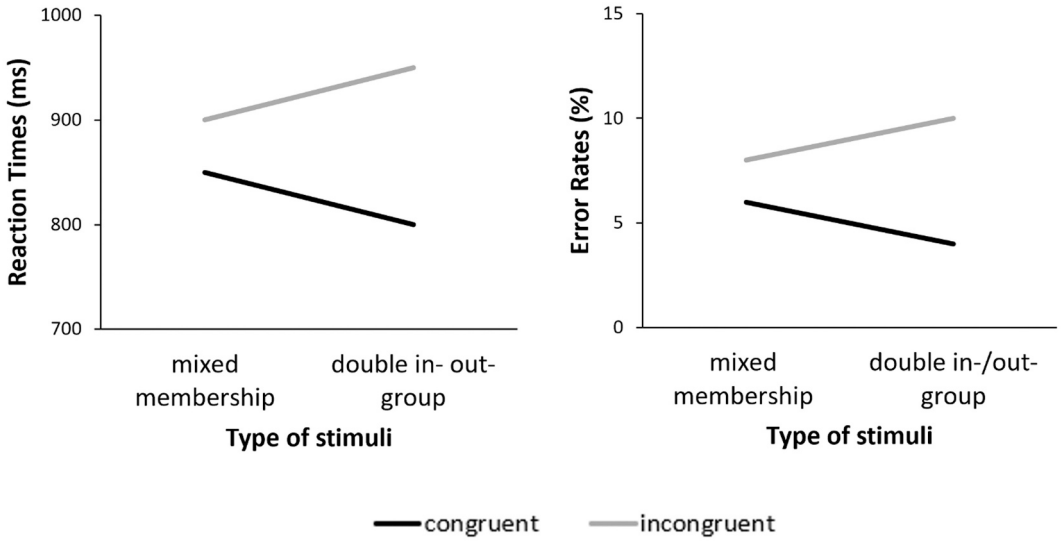

Fig. A9. Expected result patterns for H2.

Appendix G. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2022.104309.

References

Abdelal, R., Herrera, Y. M., Johnston, A. I., & McDermott, R. (2009). Introduction. In R. Abdelal, Y. M. Herrera, A. I. Johnston, & R. McDermott (Eds.), *Measuring identity: A guide for social scientists*. Cambridge: Cambridge University Press.

Abrams, D. (1994). Social self-regulation. *Personality and Social Psychology Bulletin, 20*(5), 473–483. https://doi.org/10.1177/0146167294205004

Abrams, D., & Hogg, M. (1990). An introduction to the social identity approach. In D. Abrams, & M. Hogg (Eds.), *Social identity theory: Constructive and critical advances* (pp. 1–9). Hemel Hempstead, UK: Harvester Wheatsheaf.

Abrams, D., & Hogg, M. (2004). Metatheory: Lessons from social identity research. *Personality and Social Psychology Review, 8*(2), 98–106. https://doi.org/10.1207/s15327957pspr0802_2

Allport, A., & Wylie, G. (1999). Task-switching: Positive and negative priming of task-set. In G. W. Humphreys, J. Duncan, & A. Treisman (Eds.), *Attention, space, and action: Studies in cognitive neuroscience* (pp. 273–296). Oxford University Press.

Amiot, C. E., de la Sablonnière, R., Terry, D. J., & Smith, J. R. (2007). Integration of social identities in the self: Toward a cognitive-developmental model. *Personality and Social Psychology Review, 11*(4), 364–388. https://doi.org/10.1177/1088868307304091

Bainbridge, W., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General, 142*(4), 1323–1334. https://doi.org/10.1037/a0033872

Benet-Martínez, V., & Haritatos, J. (2005). Bicultural identity integration (BII): Components and psychosocial antecedents. *Journal of Personality, 73*(4), 1015–1050. https://doi.org/10.1111/j.1467-6494.2005.00337.x

Benjamin, D., Choi, J., & Strickland, A. (2010). Social identity and preferences. *American Economic Review, 100*(4), 1913–1928. https://doi.org/10.1257/aer.100.4.1913

Brown, R. J., & Turner, J. C. (1979). The criss-cross categorization effect in intergroup discrimination. *The British Journal of Social and Clinical Psychology, 18*, 371–383. https://doi.org/10.1111/j.2044-8260.1979.tb00907.x

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition, 2*(1). https://doi.org/10.5334/joc.72

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1). https://doi.org/10.5334/joc.10

Carpenter, T., Pogacar, R., Pullig, C., Kouril, M., Aguilar, S., LaBouff, J., et al. (2019). Survey-software implicit association tests: A methodological and empirical analysis. *Behavior Research Methods, 51*(5), 2194–2208. https://doi.org/10.3758/s13428-019-01293-3

Chakravarty, S., & Fonseca, M. (2017). Discrimination via exclusion: An experiment on group identity and club goods. *Journal of Public Economic Theory, 19*(1), 244–263. https://doi.org/10.1111/jpet.12182

Crisp, R. (2010). Cross-categorization. In J. M. Levine, & M. A. Hogg (Eds.), *Encyclopedia of group processes & intergroup relations* (pp. 164–166). Thousand Oaks, CA: SAGE Publications, Inc.. https://doi.org/10.4135/9781412972017.n51

Cruwys, T., Steffens, N. K., Haslam, S. A., Haslam, C., Jetten, J., & Dingle, G. A. (2016). Social identity mapping: A procedure for visual representation and assessment of subjective multiple group memberships. *The British Journal of Social Psychology, 55*(4), 613–642. https://doi.org/10.1111/bjso.12155

Deaux, K., Reid, A., Mizrahi, K., & Ethier, K. A. (1995). Parameters of social identity. *Journal of Personality and Social Psychology, 68*(2), 280–291. https://doi.org/10.1037/0022-3514.68.2.280

Deschamps, J.-C., & Doise, W. (1978). Crossed category membership in intergroup relations. In H. Tajfel (Ed.), *Differentiation between social groups: Studies in the social psychology of intergroup relations*. London: Academic Press.

Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology, 31*(5), 410–436. https://doi.org/10.1006/jesp.1995.1018

Elchlepp, H., Best, M., Lavric, A., & Monsell, S. (2017). Shifting attention between visual dimensions as a source of switch costs. *Psychological Science, 28*(4), 470–481. https://doi.org/10.1177/0956797616686855

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the implicit association tests. *Basic and Applied Social Psychology, 27*(4), 307–316. https://doi.org/10.1207/s15324834basp2704_3

Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology, 57*(2), 239–249. https://doi.org/10.1037/0022-3514.57.2.239

Gocłowska, M. A., & Crisp, R. J. (2014). How dual-identity processes foster creativity. *Review of General Psychology, 18*(3), 216 236. https://doi.org/10.1037/gpr0000008

Graham, B., & Lavric, A. (2021). Preparing to switch languages versus preparing to switch tasks: Which is more effective? *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001027. Advance online publication.

Greenwald, A., McGhee, D., & Schwartz, J. (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., Nosek, B., & Banaji, M. (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

Greenwald, A. G., Brendl, M., Cai, H., Cvencek, D., Dovidio, J. F., Friese, M., … Wiers, R. (2020, April 7). *The implicit association test at age 20: What is known and what is not known about implicit bias*. https://doi.org/10.31234/osf.io/bf97c

Hackel, L., Coppin, G., Wohl, M., & Van Bavel, J. (2018). From groups to grits: Social identity shapes evaluations of food pleasantness. *Journal of Experimental Social Psychology, 74*, 270–280. https://doi.org/10.1016/j.jesp.2017.09.007

Haslam, C., Jetten, J., Cruwys, T., Dingle, G. A., & Haslam, S. A. (2018). *The new psychology of health: Unlocking the social cure*. New York: Routledge.

Haslam, S. A., Jetten, J., Postmes, T., & Haslam, C. (2009). Social identity, health and well-being: An emerging agenda for applied psychology. *Applied Psychology, 58*(1), 1–23. https://doi.org/10.1111/j.1464-0597.2008.00379.x

Haslam, S. A., Oakes, P. J., Reynolds, K. J., & Turner, J. C. (1999). Social identity salience and the emergence of stereotype consensus. *Personality and Social Psychology Bulletin, 25*(7), 809–818. https://doi.org/10.1177/0146167299025007004

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*(10), 1369–1385. https://doi.org/10.1177/0146167205275613

Hogg, M., & Terry, D. (2000). Social identity and self-categorization processes in organizational contexts. *Academy of Management Review, 25*(1), 121–140. https://doi.org/10.5465/amr.2000.2791606

Kharkhurin, A. V., & Wei, L. (2015). The role of code-switching in bilingual creativity. *International Journal of Bilingual Education and Bilingualism, 18*(2), 153–169. https://doi.org/10.1080/13670050.2014.884211

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching–a review. *Psychological Bulletin, 136*(5), 849–874. https://doi.org/10.1037/a0019842

Koschate, M., Naserian, E., Dickens, L., Stuart, A., Russo, A., & Levine, M. (2021). ASIA: Automated social identity assessment using linguistic style. *Behavior Research Methods, 53*(4), 1762–1781. https://doi.org/10.3758/s13428-020-01511-3

Lai, C., Skinner, A., Cooley, E., Murrar, S., Brauer, M., Devos, T., et al. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General, 145*(8), 1001–1016. https://doi.org/10.1037/xge0000179

Lavric, A., Mizon, G. A., & Monsell, S. (2008). Neurophysiological signature of effective anticipatory task-set control: A task-switching investigation. *European Journal of Neuroscience, 28*(5), 1016–1029. https://doi.org/10.1111/j.1460-9568.2008.06372.x

Longman, C., Lavric, A., & Monsell, S. (2013). More attention to attention? An eye-tracking investigation of selection of perceptual attributes during a task switch. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(4), 1142–1151. https://doi.org/10.1037/a0030409

Longman, C., Lavric, A., Munteanu, C., & Monsell, S. (2014). Attentional inertia and delayed orienting of spatial attention in task-switching. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1580–1602. https://doi.org/10.1037/a0036552

Luna, D., Ringberg, T., & Peracchio, L. (2008). One individual, two identities: Frame switching among biculturals. *Journal of Consumer Research, 35*(2), 279–293. https://doi.org/10.1086/586914

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1423–1442. https://doi.org/10.1037/0278-7393.22.6.1423

Meiran, N., & Marciano, H. (2002). Limitations in advance task preparation: Switching the relevant stimulus dimension in speeded same—Different comparisons. *Memory & Cognition, 30*(4), 540–550. https://doi.org/10.3758/bf03194955

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*, 25–40. https://doi.org/10.1006/jmla.1998.2602

Minear, M., & Park, D. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36*(4), 630–633. https://doi.org/10.3758/bf03206543

Molenberghs, P., & Louis, W. R. (2018). Insights from fMRI studies into ingroup bias. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01868. Article 1868.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134–140. https://doi.org/10.1016/s1364-6613(03)00028-7

Monsell, S. (2015). Task-set control and task switching. In J. M. Fawcett, E. F. Risko, & A. Kingstone (Eds.), *The handbook of attention* (pp. 139–172). MIT Press.

Monsell, S., & Mizon, G. (2006). Can the task-cuing paradigm measure an endogenous task-set reconfiguration process? *Journal of Experimental Psychology: Human Perception and Performance, 32*(3), 493–516. https://doi.org/10.1037/0096-1523.32.3.493

Nosek, B., Banaji, M., & Greenwald, A. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*(1), 101–115. https://doi.org/10.1037/1089-2699.6.1.101

Onorato, R., & Turner, J. (2004). Fluidity in the self-concept: The shift from personal to social identity. *European Journal of Social Psychology, 34*(3), 257–278. https://doi.org/10.1002/ejsp.195

Otten, S., & Moskowitz, G. B. (2000). Evidence for implicit evaluative in-group bias: Affect-biased spontaneous trait inference in a minimal group paradigm. *Journal of Experimental Social Psychology, 36*(1), 77–89. https://doi.org/10.1006/jesp.1999.1399

Otten, S., & Wentura, D. (1999). About the impact of automaticity in the Minimal Group Paradigm: Evidence from affective priming tasks. *European Journal of Social Psychology, 29*(8), 1049–1071. https://doi.org/10.1002/(SICI)1099-0992(199912)29:8<1049::AID-EJSP985>3.0.CO;2-Q

Pinter, B., & Greenwald, A. (2010). A comparison of minimal group induction procedures. *Group Processes & Intergroup Relations, 14*(1), 81–98. https://doi.org/10.1177/1368430210375251

Rogers, R., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Röhner, J., Schröder-Abé, M., & Schütz, A. (2013). What do fakers actually do to fake the IAT? An investigation of faking strategies under different faking conditions. *Journal of Research in Personality, 47*(4), 330–338. https://doi.org/10.1016/j.jrp.2013.02.009

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime (version 2.0). [computer software and manual]*. Pittsburgh, PA: Psychology Software Tools Inc.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*(2), 149–178. https://doi.org/10.1002/ejsp.2420010202

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole.

Turner, J., & Oakes, P. (1986). The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology, 25*(3), 237–252. https://doi.org/10.1111/j.2044-8309.1986.tb00732.x

Turner, J., Oakes, P., Haslam, S., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin, 20*(5), 454–463. https://doi.org/10.1177/0146167294205002

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

Verkuyten, M., & Hagendoorn, L. (1998). Prejudice and self-categorization: The variable role of authoritarianism and in-group stereotypes. *Personality and Social Psychology Bulletin, 24*(1), 99–110. https://doi.org/10.1177/0146167298241008

Voci, A. (2006). The link between identification and in-group favouritism: Effects of threat to social identity and trust-related emotions. *British Journal of Social Psychology, 45*, 265–284. https://doi.org/10.1348/014466605X52245

Warriner, A., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology, 46*(4), 361–413. https://doi.org/10.1016/S0010-0285(02)00520-0

Waszak, F., Hommel, B., & Allport, A. (2004). Semantic generalization of stimulus-task bindings. *Psychonomic Bulletin & Review, 11*(6), 1027–1033. https://doi.org/10.3758/BF03196732

Waszak, F., Hommel, B., & Allport, A. (2005). Interaction of task readiness and automatic retrieval in task switching: Negative priming and competitor priming. *Memory & Cognition, 33*(4), 595–610. https://doi.org/10.3758/BF03195327

Xiao, Y., & Van Bavel, J. (2019). Sudden shifts in social identity swiftly shape implicit evaluation. *Journal of Experimental Social Psychology, 83*, 55–69. https://doi.org/10.1016/j.jesp.2019.03.005