Group Processes & Intergroup Relations
G P I R

Article

A case of mistaken identity: Miscategorisation of the ingroup as a historically rivalrous outgroup triggers collective narcissism

Group Processes & Intergroup Relations
2026, Vol. 29(1) 78–96

DOI: 10.1177/13684302251345405
journals.sagepub.com/home/gpi
Sage

Rita Guerra,[1] Agnieszka Golec de Zavala,[2] Kinga Bierwiaczonek,[1,3] Pawel Ciesielski,[4] Georgios Abakoumkin,[5] Tim Wildschut[6] and Constantine Sedikides[6]

Abstract

Collective narcissism's links with intergroup relations, such as intergroup hostility, are well established, but less is known about the intergroup conditions that trigger it. We experimentally examined whether categorisation threat—operationalised as mistaking the ingroup for a historically rivalrous outgroup, thus undermining the ingroup's uniqueness—heightens collective narcissism, and whether this, in turn, escalates hostility toward the pertinent outgroup through collective narcissism. Additionally, we compared collective narcissism to another form of ingroup positivity: ingroup satisfaction. We conducted four experiments (N = 1,537) manipulating categorisation threat in two national contexts (Poland, Portugal), and carried out an internal meta-analysis. As hypothesised, the findings revealed an increase in collective narcissism, as well as a positive indirect effect of categorisation threat on outgroup hostility mediated by collective narcissism, but not by ingroup satisfaction. This research establishes categorisation threat as a robust trigger of collective narcissism.

Keywords

categorisation threat, collective narcissism, ingroup distinctiveness, ingroup satisfaction, outgroup hostility

Paper received 11 April 2024; revised version accepted 5 May 2025.

[1]Instituto Universitário de Lisboa, Portugal
[2]University of London, UK
[3]University of Oslo, Norway
[4]University of Social Sciences and Humanities, Poland
[5]University of Thessaly, Greece
[6]University of Southampton, UK

Corresponding author:
Rita Guerra, Centro de Investigação e Intervenção Social, Iscte – Instituto Universitário de Lisboa (CIS_Iscte), Av. das Forças Armadas, 40 Edifício 4, Lisboa 1649-026, Portugal.
Email: ana_rita_guerra@iscte-iul.pt

Collective narcissism, a belief that the ingroup's exceptionality is not sufficiently recognised by others, is associated with intergroup hostility, prejudice, conspiratorial thinking, and political extremism (Golec de Zavala, 2023; Golec de Zavala et al., 2019). Although the consequences of collective narcissism are well established, less is known about its situational triggers. One research stream has focused on individual-level predictors of it, such as low personal control (Cichocka et al., 2018; Marchlewska et al., 2020), individual narcissism (Golec de Zavala et al., 2019), low self-esteem (Golec de Zavala et al., 2020), high attachment anxiety (Marchlewska et al., 2022), and low political knowledge (Michalski et al., 2023). Another research stream has focused on intergroup predictors of collective narcissism, such as subjective long-term disadvantage of the ingroup (Marchlewska et al., 2018), perceived intergroup threats (Guerra et al., 2022), and social identity threats (Bagci et al., 2023; Guerra et al., 2022). Despite substantial evidence linking collective narcissism to factors at both the individual and group levels, experimental evidence identifying its triggers is scarce (cf. Bertin et al., 2022).

We addressed this knowledge gap. Building on recent intergroup approaches to collective narcissism (Guerra et al., 2022), we examined the triggering potential of categorisation threat arising from miscategorisation that undermines an ingroup's distinctiveness (Barreto et al., 2010). We operationalised categorisation threat as being mistaken for a historically rivalrous national outgroup and not being treated on the basis of one's national distinctive identity. We hypothesised that categorisation threat heightens collective narcissism and, indirectly, exacerbates hostility toward the outgroup responsible for the categorisation threat (i.e., the group with which the ingroup is conflated), with this effect being mediated through collective narcissism. We tested these hypotheses in four experiments across two national contexts (Poland, Portugal). We were concerned with national collective narcissism, although, for brevity, we refer to it as collective narcissism.

Social Identity Threats

Social identity threats refer to threats "experienced at the social identity level" (Branscombe et al., 1999, p. 36) resulting from any situation where core identity motives are undermined by the social context (Breakwell, 1986; Vignoles et al., 2000, 2006). These threats can refer to values, distinctiveness, or categorisation. Value threat (e.g., being discriminated against as a group member) entails reduced mental health, and this effect is stronger for pervasive (vs. single-event) discrimination (Emmer et al., 2024; Schmitt et al., 2014). Ingroup value threat (e.g., unfavourable competence- or morality-based judgments directed at the ingroup) evokes negative emotions (e.g., anger) and hostility (Ellemers et al., 2002). Distinctiveness threat, the sense that the ingroup is not sufficiently different from an outgroup, engenders negative emotions (e.g., hatred, disgust) toward the threatening outgroup and increases ingroup bias (Leonardelli et al., 2010) and intergroup differentiation (e.g., discrimination; Ellemers et al., 2002) in an effort to restore distinctiveness (Branscombe et al., 1999; Jetten et al., 2004). Finally, categorisation threat, being categorised against one's will as a member of a different group (Branscombe et al., 1999), instigates negative self-directed affect and lower personal self-esteem (Barreto et al., 2010) as well as distancing from the disagreeable group (Barreto & Ellemers, 2003, 2010; Ellemers et al., 2002).

Categorisation can be a source of threat for various reasons. Specifically, one may regard group membership as incongruous or unsuitable within a particular context (Branscombe et al., 1999), may not identify strongly with the group, or may prefer to be treated based on either a less or more distinctive identity (Barreto et al., 2010; Ellemers et al., 2002). Thus, categorisation threat is likely to emerge from the frustrated desire for uniqueness (Barreto et al., 2010; Ellemers et al., 2002). Although categorisation threat may seem similar to distinctiveness threat, the two are conceptually and empirically different (Barreto & Ellemers, 2002; Barreto et al., 2010). Analogous to distinctiveness threat, categorisation threat can

arise from the undermining of ingroup distinctiveness, such as when individuals seek recognition based on a unique identity but are instead treated in a manner that disregards this identity. In contrast to distinctiveness threat, categorisation threat can also arise from the desire to be recognised and treated in accordance with a less distinctive identity (Barreto et al., 2010). Moreover, categorisation threat differs from distinctiveness threat in that it is externally rather than internally induced. Whereas distinctiveness threat stems from an individual realising that their ingroup is too similar to a relevant outgroup, categorisation threat stems from the failure of others to recognise the individual's unique and preferred self-categorisation. Hence, at the core of categorisation threat is "the very fact that people's preferred self-categorisations do not correspond to the way they are perceived by others" (Branscombe et al., 1999, p. 38). The divergence between an individual's preferred self-categorisation and the categorisation imposed by others may be either contextual, referring to a self-definition preferred in a specific context, or chronic, indicating a persistent miscategorisation based on salient characteristics such as gender or race (Barreto et al., 2010; Branscombe et al., 1999).

Most studies examining the impact of categorisation threat have not considered the intentionality of the source of the miscategorisation. In their taxonomy of social identity threat, Branscombe et al. (1999) address the detrimental effects of this form of threat and refer to involuntary categorisation, implying that misategorisation may occur inadvertently. Consistent with this reasoning, a review examining the interplay between internal and external social identities posited that external categorisations are predominantly influenced by the cognitive salience of category membership, which is elicited by readily accessible visible cues (e.g., race, ethnicity, gender) or by the numerical distinctiveness of group membership (Barreto & Ellemers, 2003). Here, we concentrate on the effects of categorisation threat that stem from a compromised desire for uniqueness, specifically in the context of being mistakenly identified as a historically rivalrous national outgroup, without addressing the intentionality behind the source of the inappropriate categorisation.

Finally, the majority of the existing literature has primarily examined the effects of categorisation threat on the self, specifically in relation to identity management strategies and identification with externally imposed categories (Barreto & Ellemers, 2003; Barreto et al., 2010). In contrast, our focus is on the influence of miscategorisation on intergroup relations.

Miscategorisation, collective narcissism, and intergroup hostility. Categorisation threat can influence intergroup relations. In a study conducted in the context of the Catholic–Protestant conflict in Northern Ireland, perceptions that others do not differentiate Catholics from Protestants were associated with higher intergroup bias and decreased tolerance on the part of both groups (Schmid et al., 2009). Asians and Latinos in the US who were asked to recall an episode where they were miscategorised as belonging to a different national group (e.g., Mexican vs. El Salvador) reported more negative perceptions toward the person who neglected their national identity and also toward the person's ethnic outgroup as a whole (Flores & Huo, 2012, Study 2). Similarly, Dominican Americans who were mistakenly categorised as African Americans reported less positive attitudes toward the outgroup for which they were mistaken (Wiley, 2019).

We specifically investigate categorisation threat that arises from the miscategorisation of the ingroup as a rivalrous outgroup by external parties, which undermines the ingroup's sense of uniqueness. Individuals rely on others to help define who they are and are motivated to sustain the positivity of both their personal and collective selves (Ellemers et al., 2002; Sedikides, 2021b). Categorisation threat disrupts this positivity. The miscategorisation of the ingroup as an outgroup undermines the ingroup's sense of uniqueness and is consequently perceived as a source of threat. We propose that this threat triggers collective narcissism, characterised by a reliance on external recognition of the ingroup's

significance and exceptionalism as a defining feature (Golec de Zavala et al., 2009, 2019).

Our proposal aligns with Fromm's (1947) view that narcissistic identity formation is a reaction to a lack of positive recognition of the self by others. It also aligns with results pointing to collective narcissists' dependence on external appreciation of the ingroup's importance: collective narcissists are preoccupied with others criticising their ingroup (Golec de Zavala et al., 2013, 2016), ignoring or excluding it (Golec de Zavala et al., 2020; Hase et al., 2021), conspiring against it (Cichocka et al., 2022; Golec de Zavala et al., 2022), failing to appreciate its worth (Bertin et al., 2022), being hostile toward it or jealous of it (Dyduch-Hazar et al., 2019), and not telling it apart from other groups (Guerra et al., 2022). Building on this evidence, we hypothesise that categorisation threat activates collective narcissism, which subsequently fosters hostility toward the outgroup for which one was mistaken (Wiley, 2019). The hostility may take the form of devaluing the outgroup, thus allowing to reestablish the ingroup's importance and positivity.

Overview

In four experiments, we tested whether categorisation threat—arising from miscategorisation that undermines the ingroup's distinctiveness—increases collective narcissism (Hypothesis 1 [H1]) and hostility against the outgroup for which the ingroup is mistaken (H2) via collective narcissism (H3). Additionally, we tested H1–H3 by examining whether categorisation threat affects specifically collective narcissism versus an alternative form of ingroup positivity, ingroup satisfaction. This construct refers to the belief that one's membership in the ingroup is valuable and a reason to be proud (Leach et al., 2008). In particular, we contrasted, in all experiments, collective narcissism with ingroup satisfaction. We did so because the two forms of ingroup positivity are usually positively associated, although they represent divergent beliefs about one's ingroup and relate differently to outgroup hostility (Golec de Zavala & Lantos, 2020; Golec de Zavala et al., 2020). To date, research on categorisation threat has not differentiated between various forms of ingroup positivity nor has it examined how such threats influence collective narcissism in comparison to ingroup satisfaction. Thus, we explore the effects of categorisation threat on these two forms of ingroup positivity without offering a directional hypothesis.

We carried out these experiments in two countries (Poland, Portugal) to examine whether the findings generalise across national contexts. As target groups for mistaken categorisation, we selected outgroups with which these nations share geographical proximity and a history of rivalry: Spain in the case of Portugal (Guerra et al., 2022), Russia in the case of Poland (Lisiakiewicz, 2018).[1] Portugal and Poland are both geographically smaller than their bordering, rivalrous countries, and historically less powerful. The pairs of countries exhibit both similarities (e.g., territorial disputes, loss of sovereignty, cultural ties) and notable differences. For instance, Portugal and Spain have maintained a stable collaborative relationship as European Union members for over 3 decades, whereas Poland and Russia have experienced a longer and more contentious relationship marked by uneven historical dominance. Although our focus is on natural groups with a history of rivalry, we acknowledge that categorisation threat can also manifest between nonrivalrous groups. Indeed, research has documented detrimental effects of this form of threat for both natural groups (e.g., gender) and artificial groups (e.g., deductive vs. inductive thinkers) that lack any history of rivalry (Barreto & Ellemers, 2003).

In Experiments 1–2, we examined whether having others mistake Portugal for Spain increases Portuguese collective narcissism (H1) and hostility against Spaniards (H2) via collective narcissism (H3). In Experiment 3, we tested the generalisability of the findings in another national context, examining whether Poland being mistaken for Russia increases Polish collective narcissism (H1) and hostility against Russians (H2) indirectly via collective narcissism (H3). In preregistered Experiment 4 (see https://osf.io/

zq6y8), we additionally contrasted categorisation threat with distinctiveness threat (i.e., emphasising similarities between Poland and Russia). In recent research, distinctiveness threat (i.e., similarity with relevant outgroups) was positively associated with collective narcissism (Guerra et al., 2022). Thus, we explored potential differences between categorisation threat and distinctiveness threat in their impact on collective narcissism, without offering a directional hypothesis. We administered all measures in random order, and randomised the order of items within each measure, separately for each participant, following up with demographic questions. In suspicion checks, no participant correctly guessed the purpose of the experiments.

Finally, external recognition is a core feature of both collective narcissism (Golec de Zavala et al., 2009, 2019) and categorisation threat. Striving for positive external recognition of one's personal or collective exceptionality is a defining feature of individual narcissism (Freis, 2018; Sedikides, 2021a) and collective narcissism (Golec de Zavala et al., 2009, 2019), respectively. Collective narcissism's contingency on external recognition of the ingroup has been empirically illustrated (Golec de Zavala, 2023). External recognition of one's ingroup is also a core element for categorisation threat, which stems from the mismatch between external or other-imposed categorisation and internal or self-categorisation (Barreto & Ellemers, 2003). To find out if these two constructs are empirically distinct, we conducted a confirmatory factor analysis (CFA) comparing a two-factor solution (collective narcissism, categorisation threat) with a one-factor solution (all items loading on the same factor). In all experiments, the two-factor solution fitted the data well and significantly better than the one-factor solution (see Table S1, Supplemental Material).

Experiment 1

Experiment 1 was a preliminary test of the hypotheses in a Portuguese sample, with the outgroup being "Spaniards."

Method

Participants. To determine the sample size, we conducted Monte Carlo power analyses with the R package "bmem" (Zhang, 2014) for a mediation model with two conditions, two correlated mediators (collective narcissism, ingroup satisfaction), and one outcome variable. We used effect sizes from prior research (Guerra et al., 2022, Study 2): $r_a = .25$ for the association between categorisation threat and collective narcissism; $r_{b1} = .32$ for the association between collective narcissism and outgroup hostility; and $r_{c1} = .18$ for the association between categorisation threat and outgroup hostility. The analyses indicated that 160 participants were needed to detect an indirect effect of categorisation threat on outgroup hostility via collective narcissism with power of .80.[2]

Although we targeted a sample size of 160, we were unable to reach it. In accordance with our stopping rule, we ended data collection at the end of the academic semester. The final sample consisted of 141 Portuguese citizens (110 women, 30 men, one preferred not to answer). Participants' age varied from 18 to 61 years ($M_{age} = 29.64$, $SD_{age} = 11.27$). A sensitivity power analysis indicated that the sample size was adequate to detect effects equal to $d = 0.47$ with .80 power. In regard to educational attainment, 42.6% of participants had a high school degree, 31.2% an undergraduate university degree, and 23.4% a postgraduate university degree.

Procedure. We recruited participants via academic networks (i.e., mailing lists of several public universities) and social media (i.e., Facebook) for an experiment ostensibly assessing beliefs about the European Union. One hundred twenty-three participants (54 employed community members, 53 undergraduate students, 12 unemployed community members, four undeclared) completed the experiment for a chance to win a €150 voucher, and 18 undergraduate students completed it in the laboratory for course credit. We collected all data via Qualtrics.

We randomly assigned participants to the categorisation threat ($n = 72$) or control ($n = 69$)

condition. Participants in both conditions watched a brief (90 s) video presented as an attention task (see Supplemental Material). In the categorisation threat condition, participants watched a video about foreigners confusing Portugal with Spain (e.g., Whitney Houston, during a concert in Lisbon, addressing the audience with "Hi Spain!"). In the control condition, participants watched a neutral video about the health benefits of turmeric: the video did not make the intergroup context salient (Bertin et al., 2022). Next, all participants responded to a manipulation check, followed by measures of collective narcissism, ingroup satisfaction, and outgroup hostility.

Measures. We used Portuguese versions of all measures, validated in prior studies (Guerra et al., 2022). Response options ranged from 1 to 7, with higher values reflecting higher levels of all variables. The manipulation check comprised four items assessing concern that the ingroup was mistaken for an outgroup (e.g., "It annoys me when others see Portuguese and Spaniards as the same"; $\alpha = .87$, $M = 5.26$, $SD = 1.31$; Guerra et al., 2022; Schmid et al., 2009). We assessed collective narcissism with a five-item scale adapted from Golec de Zavala et al. (2009; e.g., "The Portuguese deserve special treatment"; $\alpha = .82$, $M = 4.16$, $SD = 1.18$). We assessed ingroup satisfaction with four items derived from Leach et al. (2008; e.g., "I'm glad to be Portuguese"; $\alpha = .90$, $M = 5.51$, $SD = 1.09$). We constructed an outgroup hostility index by combining two scales: negative emotions against Spaniards, assessed with a seven-item scale adapted from Cottrell and Neuberg (2005; e.g., "When thinking of Spaniards, I feel angry"; $\alpha = .87$, $M = 1.49$, $SD = 0.75$), and hostile behavioural intentions toward Spaniards, assessed with a 10-item scale adapted from Mackie et al. (2000; e.g., "When thinking of or interacting with Spaniards, I want to hurt them"; $\alpha = .90$, $M = 1.60$, $SD = 0.90$). The two scales were positively correlated, $r(139) = .52$, $p < .001$. Consequently, we aggregated responses to the scales into a single index of outgroup hostility.[3] We present descriptive statistics and correlations in Table 1.

Table 1. Means, standard deviations, and zero-order correlations: Experiment 1.

Measure	1	2	*M*	*SD*
1. Collective narcissism	-		4.16	1.18
2. Ingroup satisfaction	.53*	-	5.51	1.09
3. Outgroup hostility	.33*	.02	1.55	0.72

*$p < .001$.

Results and Discussion

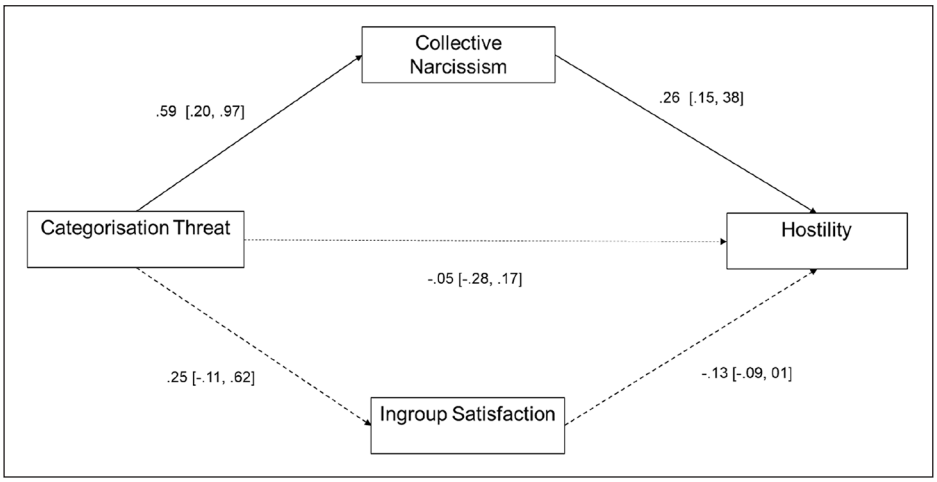
As intended, participants indicated greater concern that the ingroup was mistaken for an outgroup in the categorisation threat ($M = 5.69$, $SD = 1.45$) than in the control ($M = 4.81$, $SD = 0.99$) condition, $t(139) = 4.25$, $p < .001$, $d = 1.23$. The manipulation was effective.

To test H1, we conducted independent sample *t* tests with collective narcissism and ingroup satisfaction as dependent variables. Consistent with H1, collective narcissism was higher in the categorisation threat ($M = 4.43$, $SD = 1.20$) than in the control ($M = 3.88$, $SD = 1.08$) condition, $t(139) = 2.87$, $p = .005$, Cohen's $d = 0.48$, 95% CI [0.15, 0.81]. Ingroup satisfaction did not differ between the categorisation threat ($M = 5.63$, $SD = 1.15$) and control ($M = 5.37$, $SD = 1.01$) conditions, $t(137) = 1.38$, $p = .169$, Cohen's $d = 0.23$, 95% CI [−0.10, 0.57].

To test H2 and H3, we conducted mediation in the multiple regression context using the PROCESS macro for SPSS (Model 4; see Figure 1; Hayes, 2018). We tested multiple mediation with condition as predictor (dummy-coded: 0 = control, 1 = categorisation threat), collective narcissism and ingroup satisfaction as parallel mediators, and the index of outgroup hostility as outcome. We used bootstrapping with 5,000 samples and 95% bias corrected confidence intervals to assess indirect effects.

In accord with H2, the total effect of condition on outgroup hostility was positive and significant, $b = 0.12$, $SE = 0.06$, 95% CI [0.03, 0.25]. In accord with H3, we obtained a positive and significant indirect effect of categorisation threat (vs. control) via collective narcissism on outgroup hostility, $b = 0.15$, $SE = 0.07$, 95% CI [0.04, 0.30]; categorisation threat increased

Figure 1. Indirect effects of categorisation threat on hostility: Experiment 1.



Note. Unstandardised regression coefficients are reported. The coefficient between independent and dependent variables is a direct effect in the presence of the mediators. Categorisation threat: 0 = control, 1 = categorisation threat.

collective narcissism, which in turn predicted higher outgroup hostility (see Figure 1). The indirect effect via ingroup satisfaction was not significant, $b = -0.03$, $SE = 0.03$, 95% CI [−0.10, 0.01].

Experiment 2

Given that Experiment 1 was underpowered, we tested its replicability in Experiment 2. Again, the sample was Portuguese, and the outgroup was "Spaniards."

Method

Participants. We conservatively oversampled to achieve high statistical power (including that for the indirect effect via ingroup satisfaction) based on identical power calculations. We tested 609 Portuguese citizens (417 women, 186 men, four preferred not to answer, two undeclared), ranging in age from 18 to 86 years ($M_{age} = 31.62$, $SD_{age} = 14.82$). Of the participants, 5.1% had less than high school education, 58.4% had a high school degree, 24.5% had an undergraduate university degree, and 11.8% had a postgraduate university degree. As per an a priori decision, we excluded an additional 90 participants who failed the attention check (a grid screener): "Please do not answer to this item. Do not select any option of the 1–7 scale. This question aims at detecting random answers." The percentage of participants who failed this attention check (13%) was lower than the typical passage rate (61 to 91%) for grid screeners (Berinsky et al., 2021).

Procedure. We recruited participants via social media (i.e., Facebook, WhatsApp) for an experiment that ostensibly assessed beliefs about the European Union, and collected all data via Qualtrics. We randomly assigned participants to the categorisation threat ($n = 292$) or control ($n = 317$) condition. The manipulations, materials, and measures were the same as in Experiment 1. Participants responded to the manipulation check ($\alpha = .83$, $M = 5.01$, $SD = 1.30$), followed by measures of collective narcissism ($\alpha = .83$, $M = 3.98$, $SD = 1.17$), ingroup satisfaction ($\alpha = .86$, $M = 5.58$, $SD = 1.03$), and outgroup hostility (negative emotions against Spaniards: $\alpha = .80$, $M = 1.33$, $SD = 0.68$; hostile behavioural intentions toward Spaniards: $\alpha = .87$, $M = 1.45$, $SD = 0.80$). As in Experiment 1, the negative emotions and hostile behavioural intentions scales were

Table 2. Means, standard deviations, and zero-order correlations: Experiment 2.

Measure	1	2	3	*M*	*SD*
1. Collective narcissism	-			3.98	1.16
2. Ingroup satisfaction	.43*	-		5.58	1.03
3. Outgroup hostility	.31*	.05	.14*	1.39	0.68

*$p < .001$.

positively correlated, $r(607) = .68$, $p < .001$. Hence, we aggregated responses to the two scales into an outgroup hostility index.[4] We report descriptive statistics and correlations in Table 2.

Results and Discussion

Participants expressed greater concern that the ingroup was mistaken for another group in the categorisation threat ($M = 5.29$, $SD = 1.20$) than in the control ($M = 4.76$, $SD = 1.35$) condition, $t(605) = 5.05$, $p < .001$, $d = 0.41$, 95% CI [0.57, 0.25]. The manipulation was successful. To test H1, we conducted independent sample *t* tests with collective narcissism and ingroup satisfaction as dependent variables. Consistent with H1, collective narcissism was higher in the categorisation threat ($M = 4.08$, $SD = 1.17$) than in the control ($M = 3.88$, $SD = 1.15$) condition, $t(605) = 2.12$, $p = .034$, Cohen's $d = 0.17$, 95% CI [0.32, 0.13]. Ingroup satisfaction was also higher in the categorisation threat ($M = 5.68$, $SD = 1.05$) than in the control ($M = 5.48$, $SD = 1.01$) condition, $t(602) = 2.45$, $p = .015$, Cohen's $d = 0.19$, 95% CI [0.36, 0.04]. Thus, the results of Experiment 2 replicated those of Experiment 1 and additionally indicated that the expected effect was not specific to collective narcissism.

To test H2 and H3, we used the PROCESS macro for SPSS (Model 4; see Figure 2 for detailed results; Hayes, 2018). We tested multiple mediation with condition as predictor (dummy-coded: 0 = control, 1 = categorisation threat), collective narcissism and ingroup satisfaction as parallel mediators, and outgroup hostility as outcome. We used bootstrapping with 5,000 samples and 95% bias corrected confidence intervals to assess indirect effects.

Contrary to H2, the total effect of condition on outgroup hostility was not significant, $b = 0.03$, $SE = 0.02$, 95% CI [−0.01, 0.07]. Congruent with H3, we obtained a positive and significant indirect effect of categorisation threat (vs. control) via collective narcissism on outgroup hostility, $b = 0.04$, $SE = 0.02$, 95% CI [0.004, 0.09]. That is, categorisation threat increased collective narcissism, which predicted higher outgroup hostility (see Figure 2). The indirect effect via ingroup satisfaction was significant, but negative, $b = -0.02$, $SE = 0.01$, 95% CI [−0.03, −0.002]. The findings of Experiment 2 did not corroborate those of Experiment 1 with respect to H2. Nevertheless, they were consistent with the findings of Experiment 1 concerning H3.

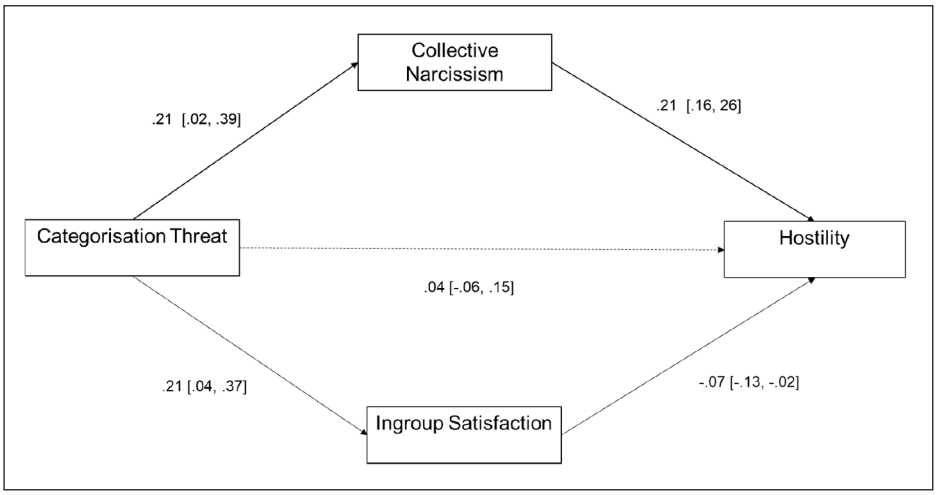
Experiment 3

In Experiment 3, the sample was Polish and the outgroup "Russians." To gauge the internal validity of the experimental manipulation and ensure that its effects were indeed due to categorisation threat rather than mere salience of the intergroup context (Brown & Hewstone, 2005), we implemented an additional control condition that referred to an intergroup context but did not contain a categorisation threat. We expected the categorisation threat condition to differ from both control conditions, such that collective narcissism would be higher in the categorisation threat versus the pooled control conditions.

Method

Participants. To determine the sample size, we conducted Monte Carlo power analyses with the R package "bmem" (Zhang, 2014) for a mediation

Figure 2. Indirect effects of categorisation threat on hostility: Experiment 2.



Note. Unstandardised regression coefficients are reported. The coefficient between independent and dependent variables is a direct effect in the presence of the mediators. Categorisation threat: 0 = control, 1 = categorisation threat.

model with three conditions, two correlated mediators, and one outcome variable. We conservatively assumed small yet theoretically meaningful effects (r = .20) for all paths to ensure that the experiment was well powered. The estimation revealed a required sample of N = 330 to detect the indirect effect of interest with power of .80. We recruited 332 Polish adults (174 women, 158 men), ranging in age from 18 to 82 years (M_{age} = 46.73, SD_{age} = 15.28), through the Ariadna Research Panel (www.panelariadna.pl). Participation of those who failed to correctly respond to attention checks was automatically discontinued.

Procedure. The experiment was ostensibly concerned with the association between attention and social attitudes. We randomly assigned participants to one of three conditions: categorisation threat (n = 112), intergroup control (n = 107), control (n = 113). All participants watched a brief video presented as an attention task (see Supplemental Material). In the categorisation threat condition, the video was about foreigners mistaking Poland for Russia (e.g., Justin Bieber, on his way to give a concert in Warsaw, Poland, expressing excitement about visiting Russia). In the intergroup control condition, the video was about bicycle lanes along the border between Poland and Russia. In the control condition, the video was about the health benefits of turmeric, as in Experiments 1–2. Next, participants responded to manipulation checks and measures of collective narcissism, ingroup satisfaction, and outgroup hostility (this time measured only via negative emotions toward Russians).

Measures. We used validated Polish versions of all measures (Golec de Zavala et al., 2023). Responses were given on a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*). Greater numbers represent higher levels of all variables. We assessed the manipulation check (α = .91, M = 4.20, SD = 1.39), collective narcissism, ingroup satisfaction, and outgroup hostility (negative emotions) as in Experiment 1. We present descriptive statistics, reliabilities, and correlations in Table 3.

Results and Discussion

To test the effectiveness of the manipulation and our hypotheses (H1, H2, H3), we computed two orthogonal contrasts: categorisation threat versus pooled intergroup control and control conditions (C1), and intergroup control versus control

Table 3. Reliabilities, means, standard deviations, and zero-order correlations: Experiment 3.

Measure	1	2	α	*M*	*SD*
1. Collective narcissism	-		.94	4.16	1.37
2. Ingroup satisfaction	.62*	-	.96	5.32	1.30
3. Outgroup hostility	.36*	.10	.85	3.02	0.99

*$p < .001$.

condition (C2). An analysis of variance (ANOVA) on the manipulation check revealed significant differences between conditions, $F(2, 329) = 12.55$, $p < .001$. Participants reported greater concern that Poland was mistaken for Russia in the categorisation threat ($M = 4.71$, $SD = 1.26$) than in the pooled control ($M = 4.06$, $SD = 1.28$) and intergroup control ($M = 3.83$, $SD = 1.49$) conditions, $t(329) = 4.88$, $p < .001$, Cohen's $d = 0.57$, 95% CI [0.33, 0.80]. The latter conditions were not significantly different, $t(329) = 1.23$, $p = .219$, Cohen's $d = 0.16$, 95% CI [−0.10. 0.43]. Together, these results suggest the manipulation was effective.

To test H1, we conducted an ANOVA on collective narcissism, $F(2, 329) = 3.99$, $p = .019$. In support of H1, orthogonal contrasts showed that collective narcissism was higher in the categorisation threat condition ($M = 4.45$, $SD = 1.40$) than in the pooled intergroup control ($M = 4.00$, $SD = 1.30$) and control ($M = 4.01$, $SD = 1.38$) conditions, $t(329) = 2.83$, $p = .005$, Cohen's $d = 0.33$, 95% CI [0.10, 0.56]. The latter two conditions did not differ significantly, $t(329) = -0.05$, $p = .962$, Cohen's $d = -0.01$, 95% CI [−0.27, 0.26]; that is, it made no difference whether the control condition included the intergroup context.

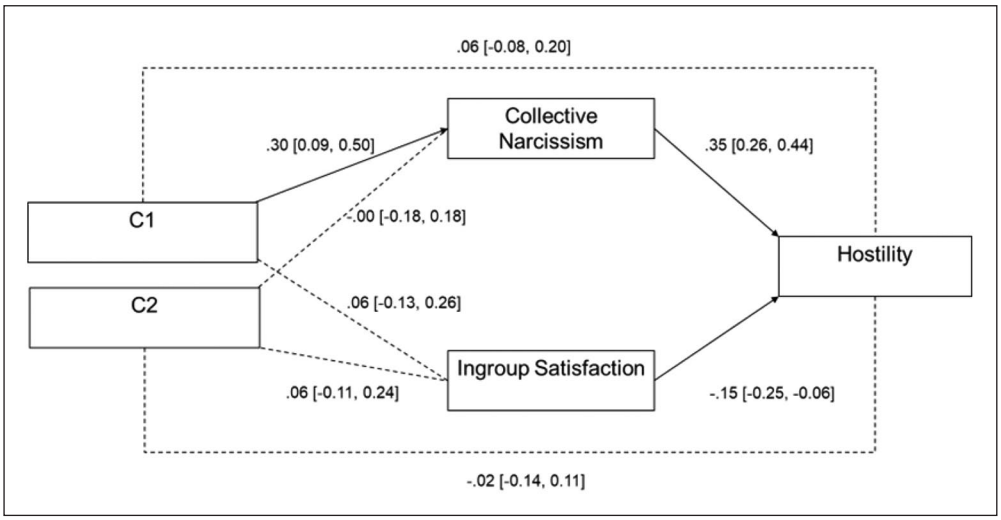
We conducted the same analysis for ingroup satisfaction. The main effect of condition was not significant, $F(2, 329) = 0.47$, $p = .627$. Ingroup satisfaction did not differ between the categorisation threat condition ($M = 5.38$, $SD = 1.36$) and the pooled intergroup control ($M = 5.35$, $SD = 1.28$) and control ($M = 5.22$, $SD = 1.25$) conditions, $t(329) = 0.63$, $p = .53$, Cohen's $d = 0.07$, 95% CI [−0.16, 0.30]. The latter two conditions did not differ significantly either, $t(329) = 0.73$, $p = .47$, Cohen's $d = 0.10$, 95% CI [−0.17, 0.36]. Taken together, as hypothesised (H1) and replicating the results of Experiments 1–2, categorisation threat increased collective narcissism.

To test H2 and H3, we conducted mediation analysis using PROCESS for SPSS (Model 4; see Figure 3; Hayes, 2018). We used bootstrapping with 5,000 samples and 95% bias corrected confidence intervals to calculate the indirect effects. We entered the experimental condition (orthogonal contrast coded: C1, C2) as predictor, collective narcissism and ingroup satisfaction as parallel mediators, and outgroup hostility as outcome. The contrast of interest compares the experimental condition with the pooled control conditions (C1 in Figure 3). Consistent with H2, and replicating the Experiment 1 results, the total effect of categorisation threat (compared to the pooled control conditions) on outgroup hostility was positive and significant, $b = 0.09$, $SE = 0.03$, 95% CI [0.03, 0.16]. In agreement with H3, and similar to Experiments 1–2, we observed a positive and significant indirect effect of categorisation threat, via collective narcissism, on outgroup hostility, $b = 0.10$, $SE = 0.04$, 95% CI [0.03, 0.18]. The indirect effect of categorisation threat on outgroup hostility via ingroup satisfaction was not significant, $b = -0.01$, $SE = 0.02$, 95% CI [−0.05, 0.02].

Experiment 4

In Experiment 4, the sample was again Polish and the outgroup "Russians." We compared categorisation threat to the intergroup control condition (as in Experiment 3), and to a distinctiveness threat condition in which we strongly emphasised

Figure 3. Indirect effects of categorisation threat on hostility: Experiment 3.



Note. Unstandardised regression coefficients are reported. Coefficients between independent variable and dependent variable are direct effects in the presence of the mediator. C1: control = −1/2, control intergroup = −1/2, categorisation threat = 1; C2: control = −1, control intergroup = 1, categorisation threat = 0.

the similarities between Poland and Russia (Jetten et al., 2004). We expected the categorisation threat condition to differ from both the intergroup control and distinctiveness threat conditions.

Method

Participants. We determined sample size using the same power calculations as in Experiment 3, but we conservatively oversampled. We tested 455 Polish adults (225 women, 230 men) ranging in age from 18 to 83 years (M_{age} = 42.57, SD_{age} = 15.99). The representative sample (with respect to gender, age, and education) was collected by the Ariadna Research Panel. As in Experiment 1, participants who failed attention checks were automatically excluded.

Procedure. We randomly assigned participants to one of three conditions (ostensibly concerned with the relation between attention and social attitudes): categorisation threat (n = 146), distinctiveness threat (n = 159), intergroup control (n = 150). Participants watched a brief video presented as an attention task (see Supplemental Material). We used the same instructions for the categorisation threat and intergroup control conditions as in Experiment 3. In the distinctiveness threat condition, the video was about similarities (e.g., language, art, architecture) between Poles and Russians. Next, all participants responded to manipulation checks and measures of collective narcissism, ingroup satisfaction, and outgroup hostility (comprised both of negative emotions and hostile behavioural intentions) against Russians.

Measures. We administered a manipulation check for categorisation threat (α = .85, M = 4.28, SD = 1.23), as in Experiment 3. The manipulation check for distinctiveness threat (i.e., similarity between Poles and Russians) comprised two items adapted from Leach et al.'s (2008) Homogeneity Subscale ("Poles and Russians are very similar to each other" and "Poles and Russians have a lot in common with each other"; α = .82, M = 4.42, SD = 1.18). We measured collective narcissism (α = .91, M = 4.25, SD = 1.30), ingroup satisfaction (α = .93, M = 5.17, SD = 1.22), and outgroup hostility (negative emotions against Russians: α = .89, M = 2.85, SD = 1.08;

Table 4. Means, standard deviations, and zero-order correlations: Experiment 4.

Measure	1	2	*M*	*SD*
Collective narcissism	-		4.25	1.30
Ingroup satisfaction	.58*	-	5.17	1.22
Outgroup hostility	.25*	.04	2.91	0.90

*$p < .001$.

hostile behavioural intentions toward Russians: $\alpha = .88$, $M = 2.97$, $SD = 0.94$) as before. As in Experiments 1–2, the negative emotions and hostile behavioural intentions scales were positively correlated, $r(453) = .58$, $p < .001$. Therefore, we aggregated responses to the two scales into an outgroup hostility index ($M = 2.91$, $SD = 0.90$).[5] We display descriptive statistics and correlations in Table 4.

Results and Discussion

We computed two orthogonal contrasts: (1) categorisation threat versus intergroup control and distinctiveness threat conditions pooled (C1), and (2) intergroup control versus distinctiveness threat condition (C2). An ANOVA on the manipulation check for categorisation threat revealed a significant omnibus effect of condition, $F(2, 452) = 33.40$, $p < .001$. Participants reported being concerned more about the possibility of Poland not being differentiated from Russia in the categorisation threat condition ($M = 4.91$, $SD = 1.18$) than in the pooled intergroup control ($M = 3.91$, $SD = 1.16$) and distinctiveness threat ($M = 4.03$, $SD = 1.11$) conditions, $t(452) = 8.13$, $p < .001$, Cohen's $d = 0.82$, 95% CI [0.61, 0.10]. The latter conditions were not significantly different, $t(452) = -0.97$, $p = .330$, Cohen's $d = -0.11$, 95% CI [−0.33, 0.11]. The manipulation of categorisation threat was effective.

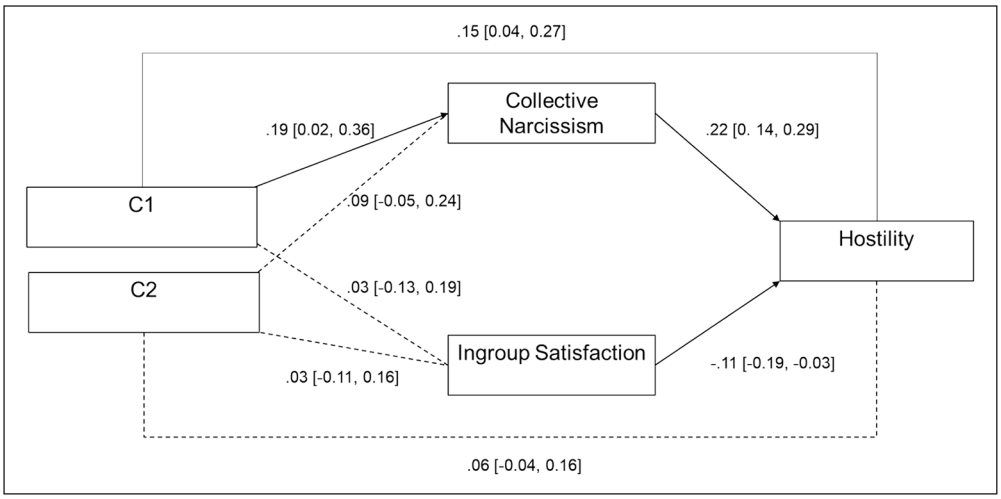
An ANOVA on the manipulation check for distinctiveness threat also revealed a significant omnibus effect of condition, $F(2, 452) = 14.15$, $p < .001$. Participants reported higher similarity between Poles and Russians in the pooled distinctiveness threat ($M = 4.64$, $SD = 1.23$) and intergroup control ($M = 4.60$, $SD = 1.04$) conditions than in the categorisation threat ($M = 4.01$, $SD = 1.14$) condition, $t(452) = 5.30$, $p < .001$, Cohen's $d = 0.53$, 95% CI [0.33, 0.73]. However, the difference between the intergroup control and distinctiveness threat conditions was not significant, $t(452) = -0.35$, $p = .731$, Cohen's $d = -0.04$, 95% CI [−0.26, 0.18]. Although the manipulation of distinctiveness threat was only partially effective, it achieved the main objective of distinguishing between categorisation threat and distinctiveness threat.

To test H1, we conducted an ANOVA on collective narcissism. The main effect of condition was significant, $F(2, 452) = 3.28$, $p = .038$. In support of H1, orthogonal contrasts showed that collective narcissism was higher in the categorisation threat condition ($M = 4.44$, $SD = 1.24$) than in the pooled intergroup control ($M = 4.05$, $SD = 1.37$) and distinctiveness threat ($M = 4.25$, $SD = 1.27$) conditions, $t(452) = 2.22$, $p = .027$, Cohen's $d = 0.44$, 95% CI [0.05, 0.84]. The intergroup control and distinctiveness threat conditions were not significantly different from each other, $t(452) = 1.32$, $p = .187$, Cohen's $d = 0.15$, 95% CI [−0.07, 0.37].

Next, we carried out the same analysis for ingroup satisfaction. The main effect of condition was not significant, $F(2, 452) = 0.14$, $p = .872$. Ingroup satisfaction did not differ between the categorisation threat condition ($M = 5.20$, $SD = 1.27$) and the pooled intergroup control ($M = 5.13$, $SD = 1.26$) and distinctiveness threat ($M = 5.18$, $SD = 1.27$) conditions, $t(452) = 0.38$, $p = .704$, Cohen's $d = 0.08$, 95% CI [−0.32, 0.47]. The latter two conditions also did not significantly differ, $t(452) = 0.37$, $p = .713$, Cohen's $d = 0.04$, 95% CI [−0.18, 0.27]. These results are in alignment with the findings of Experiments 1 and 3.

To test H2 and H3, we carried out mediation analysis using PROCESS for SPSS (Model 4; see

Figure 4. Indirect effects of categorisation threat on hostility: Experiment 4.



Note. Unstandardised regression coefficients are reported. Coefficients between independent variable and dependent variable are direct effects in the presence of the mediator. C1: intergroup control = −1/2, distinctiveness threat = −1/2, categorisation threat = 1; C2: intergroup control = −1, distinctiveness threat = 1, categorisation threat = 0.

Figure 4; Hayes, 2018). We used bootstrapping with 5,000 samples and 95% bias corrected confidence intervals to calculate the indirect effects. We entered the experimental condition (orthogonal contrast coded: C1, C2) as predictor, collective narcissism and ingroup satisfaction as parallel mediators, and outgroup hostility as outcome. The contrast of interest compares the categorisation threat condition with the pooled intergroup control and distinctiveness threat conditions (C1 in Figure 4), given that the latter two did not differ in the previous analyses. Consistent with H2, the total effect of categorisation threat (compared to pooled intergroup control and distinctiveness threat conditions) on outgroup hostility was significant, $b = 0.04$, $SE = 0.02$, 95% CI [0.01, 0.08]. Consistent with H3, we found a positive and significant indirect effect of categorisation threat, via collective narcissism, on outgroup hostility, $b = 0.04$, $SE = 0.02$, 95% CI [0.01, 0.09]. The indirect effect of categorisation threat on outgroup hostility, via ingroup satisfaction, was not significant, $b = -0.00$, $SE = 0.01$, 95% CI [−0.03, 0.01]. The indirect effects of distinctiveness threat (compared to intergroup control) on outgroup hostility, via collective narcissism ($b = 0.02$, $SE = 0.02$, 95% CI [−0.01, 0.06]) and ingroup satisfaction ($b = -0.00$, $SE = 0.01$, 95% CI [−0.02, 0.01]), as well as the total effect, were not significant. The results replicate those of Experiments 1 and 3 concerning H2, and the results of Experiments 1–3 concerning H3.

Internal Meta-Analysis

As we observed some inconsistencies across experiments, we conducted an internal meta-analysis to check the robustness of the findings.

Method

We tested the indirect effect across all experiments using the two-stage meta-analytical structural equations modeling approach (TSSEM; Cheung, 2015a, 2022). In the first stage, we meta-analysed the correlations between each pair of variables in our model (categorisation threat, collective narcissism, ingroup satisfaction, outgroup hostility) from each experiment. As in Experiments 1–4, we coded categorisation threat as 1 and the remaining pooled conditions as 0. Given a single correlation per experiment for each pair of variables, there was no interdependency between these effects, allowing for a standard random effects model.

Table 5. Meta-analytical estimates of mediation model paths.

	Estimate	*SE*	95% CI low	95% CI high	z	*p*
CT_CN	0.12	0.02	0.07	0.17	4.77	< .001
CN_H	0.39	0.03	0.33	0.44	14.29	< .001
IS_H	−0.16	0.03	−0.21	−0.10	−5.64	< .001
CT_H	0.05	0.02	0.01	0.10	2.17	0.02
CT_IS	0.06	0.02	0.01	0.11	2.48	0.01
CN_IS	0.53	0.02	0.49	0.57	29.49	< .001

CT: categorisation threat; CN: Collective narcissism; IS: Ingroup satisfaction; H: Hostility.

Thus, we converted all correlations to z scores with a Fisher *r*-to-z transformation, we pooled them in the "metafor" package for R (Viechtbauer, 2010), and converted the results back to *r*. This procedure resulted in a pooled correlation matrix which, in the second stage, we fed into the "metaSEM" package for R (Cheung, 2015b) to fit a mediation model. As bootstrapping procedures are currently unavailable for TSSEM, we additionally tested the indirect effects using the Sobel, Aroian, and Goodman tests (Obaidi et al., 2023).

Results and Discussion

We present in Table 5 the estimates of the mediation model. Given that all relations between variables were estimated, the model was fully saturated and so fit indices do not apply. Across experiments, we found a small, significant, and homogenous positive effect of categorisation threat on collective narcissism ($\tau^2 < .001$, $I^2 = 0.20\%$, 95% prediction intervals [PI] [0.07, 0.17]),[6] supporting H1; and a moderate, significant association between collective narcissism and outgroup hostility that showed low heterogeneity ($\tau^2 < .001$, $I^2 = 12.77\%$, 95% PI [0.25, 0.37]). In line with H2, the total effect of categorisation threat on outgroup hostility (obtained in the first stage of TSSEM, before including ingroup satisfaction and the indirect effects in the model), although small, was also significant ($r = .09$, 95% CI [0.03, 0.15]) with low heterogeneity ($\tau^2 < .001$, $I^2 = 19.63\%$, 95% PI [0.01, 0.16]). Similarly, we found, in the second stage, a small and significant direct effect. Finally, in accord with H3, Sobel, Aroian, and Goodman tests (Obaidi et al., 2023) supported the meta-analytical indirect effect of categorisation on outgroup hostility via collective narcissism ($B = 0.05$, all SEs $= 0.01$, all ps $< .001$).

General Discussion

We examined whether categorisation threat (i.e., others mistaking one's national ingroup for a historically rivalrous outgroup) increases collective narcissism, and whether collective narcissism in turn predicts increased hostility toward the outgroup. In line with the proposal that contingency on external recognition is a core feature of collective narcissism (Golec de Zavala et al., 2009, 2019), across four experiments, collective narcissism increased when participants were led to believe that others miscategorised the ingroup, failing to distinguish it from a national outgroup. Also, as hypothesised, across experiments, categorisation threat increased outgroup hostility via elevated collective narcissism.

The findings align with existing literature indicating that collective narcissism is linked to a concern over insufficient positive regard for the ingroup by others (Golec de Zavala et al., 2009, 2016), a lack of recognition of the ingroup's significance (Bertin et al., 2022), disregard or neglect of the ingroup (Hase et al., 2021), and failure to adequately distinguish the ingroup from other groups (Guerra et al., 2022). Moreover, we demonstrated that merely having the ingroup mistaken for a historically antagonistic outgroup was sufficient to elevate levels of collective narcissism. This

categorisation threat led to a specific increase in collective narcissism, while levels of ingroup satisfaction remained unaffected (except in Experiment 2). Stated otherwise, categorisation threat emerged as an intergroup trigger of collective narcissism, but not of other aspects of positive ingroup identification (i.e., ingroup satisfaction).

A precise understanding of the specific triggers that activate collective narcissism is crucial for developing effective interventions aimed at reducing it. Future investigations could address the conditions that may mitigate the adverse effects of categorisation threat on collective narcissism. In one study, the detrimental effects of categorisation threat on group identification and loyalty were attenuated when participants' self-chosen identities were respected (Barreto & Ellemers, 2002), suggesting that perceived respect might serve as a potential strategy for mitigating the adverse effects of categorisation threat. This suggestion aligns with research indicating that respect, particularly equality-based respect, is a predictor of positive intergroup relations (Simon & Grabow, 2014).

As mentioned above, categorisation threat increased collective narcissism but not ingroup satisfaction in all experiments but one, but it also augmented ingroup satisfaction in Experiment 2. This discrepancy might indicate the presence of moderators. Subsequent studies could explore the conditions under which categorisation threat also elevates ingroup satisfaction. Such research may inform the development of interventions aimed at reducing outgroup hostility following categorisation threat. Based on prior findings, we expect that, when categorisation threat heightens both collective narcissism and ingroup satisfaction, the impact on outgroup hostility will be attenuated due to the overlap between these constructs, which exhibit opposing unique associations with hostility toward outgroups.

Additionally, we illustrated that, although external recognition is a core conceptual feature of both collective narcissism and categorisation threat, these two constructs are theoretically and empirically distinct, given that they emerged as separate factors in CFAs across experiments.

Our research has certain limitations. First, across experiments, the effects were small. However, they were statistically significant and consistent, as indicated by the low levels of heterogeneity detected in the meta-analysis. Also, the effect sizes were similar to those found in experiments that manipulated collective narcissism (Bertin et al., 2022). Second, only the first path of our mediation model (Path A) was experimental, whereas the second path (b) was cross-sectional (Maxwell & Cole, 2007). However, we followed recent recommendations to address this criticism: we tested a theoretically driven causal model and accounted for alternative mediators (i.e., ingroup satisfaction; Fiedler et al., 2018). Although we varied the intergroup dimensions, replication across diverse intergroup contexts and cultural settings would strengthen the reliability of our findings—for example, by examining the effects of miscategorisation among groups lacking a history of rivalry. Finally, given the similarities between categorisation threat and distinctiveness threat, future research could systematically compare the specific impact of the lack of external recognition of ingroup distinctiveness with the broader effects of diminished self-perceived distinctiveness (Jetten et al., 2004; Leonardelli et al., 2010). Additionally, future studies could investigate the potentially differential effects of various forms of social identity threats (e.g., threats to group value or group distinctiveness) on collective narcissism and ingroup satisfaction, thereby further disentangling not only the consequences but also the antecedents of these two forms of ingroup positivity. To enable a more rigorous comparison of different forms of social identity threat, it may be beneficial to include different control conditions that more closely parallel the categorisation threat manipulation—for instance, by presenting the same external source correctly versus incorrectly categorising the ingroup—rather than relying on neutral, unrelated content.

Taken together, in four experiments across two national contexts, categorisation threat triggered collective narcissism and indirectly predicted hostility against the groups for which the ingroup was mistaken. These findings, bolstered by a

meta-analysis, advance the literature on collective narcissism and its antecedents. By demonstrating the role of categorisation threat arising from miscategorisation, they highlight the importance of considering not only individual-level variables (e.g., self-esteem, Golec de Zavala et al., 2019; personal control, Cichocka et al., 2018), but also the social context in which groups function when predicting collective narcissism.

Author Note

Pawel Ciesielski is currently affiliated to Faculty of Psychology and Cognitive Sciences, Adam Mickiewicz University in Poznań, Poland.

Data Availability Statement

We deposited datasets, codes for the analyses, and supplemental material at the Open Science Framework (OSF; https://osf.io/a75dj/).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval

We received ethical approval for all experiments from the first author's institution (Iscte – Instituto Universitário de Lisboa).

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a Foundation of Science and Technology (Fundação para a Ciência e Tecnologia) grant awarded to Rita Guerra, Agnieszka Golec de Zavala, and Constantine Sedikides (PTDC/MHC-PSO/0144/2014), and by a Polish National Science Centre (Narodowe Centrum Nauki) grant awarded to Agnieszka Golec de Zavala (2017/26/A/HS6/00647).

ORCID iDs

Rita Guerra https://orcid.org/0000-0003-3184-5164
Agnieszka Golec de Zavala https://orcid.org/0000-0002-7631-9486
Georgios Abakoumkin https://orcid.org/0000-0002-1671-3561

Notes

1. We collected these data before the Russian invasion of Ukraine.
2. We based our sample size decisions on the indirect effect via collective narcissism. We did not calculate power for the indirect effect via ingroup satisfaction because the relevant correlations for ingroup satisfaction were small in Guerra et al. (2022, Study 2): for ingroup satisfaction and hostility, $r = .02$; and for ingroup satisfaction and distinctiveness threat, $r = .17$. The power for the a * b product based on these correlations is .01 with 160 participants, and .28 with 5,000 participants. We calculated sensitivity power analyses based on the smallest effect that we considered theoretically meaningful ($r = .20$). With 160 participants, the power of the indirect effect via ingroup satisfaction was .42. To detect this effect with power of .80, we would need to test 330 participants. Experiment 1, then, was underpowered, an issue we addressed in Experiment 3.
3. Separate analyses for each measure produced similar results (Table S2, Supplemental Material).
4. Separate analyses for each measure produced similar results (Table S2, Supplemental Material).
5. Separate analyses for each measure produced similar results (Table S2, Supplemental Material).
6. The heterogeneity statistics provided in the text were obtained in the first stage of TSSEM; hence, the prediction intervals (PI) might not match the estimates of the path model.

References

Barreto, M., & Ellemers, N. (2002). The impact of respect versus neglect of self-identities on identification and group loyalty. *Personality and Social Psychology Bulletin*, *28*(5), 629–639. https://doi.org/10.1177/0146167202288007

Barreto, M., & Ellemers, N. (2003). The effects of being categorized: The interplay between internal and external social identities. *European Review of Social Psychology*, *14*(1), 139–170. https://doi.org/10.1080/10463280340000045

Barreto, M., Ellemers, N., Scholten, W., & Smith, H. (2010). To be or not to be: The impact of implicit versus explicit inappropriate social categorisations on the self. *British Journal of Social Psychology*, *49*(1), 43–67. https://doi.org/10.1348/014466608X400830

Berinsky, A. J., Margolis, M. F., Sances, M. W., & Warshaw, C. (2021). Using screeners to measure

respondent attention on self-administered surveys: Which items and how many? *Political Science Research and Methods*, *9*(2), 430–437. http://doi:org/10.1017/psrm.2019.53

Bertin, P., Marinthe, G., Biddlestone, M., & Delouvée, S. (2022). Investigating the identification–prejudice link through the lens of national narcissism: The role of defensive group beliefs. *Journal of Experimental Social Psychology*, *98*, Article 104252. https://doi.org/10.1016/j.jesp.2021.104252

Branscombe, N. R., Ellemers, N., & Spears, R. (1999). The context and content of social identity threats. In R. Ellemers, R. Spears, & B. Doosje (Eds.), *Social identity: Context, commitment, content* (pp. 35–58). Wiley-Blackwell.

Breakwell, G. M. (1986). *Coping with threatened identities*. Routledge.

Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. *Advances in Experimental Social Psychology*, *37*, 255–343. https://doi.org/10.1016/S0065-2601(05)37005-5

Cheung, W.-L. M. (2015a). *Meta-analysis: A structural equation modeling approach*. Wiley.

Cheung, W.-L. M. (2015b). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, *5*, Article 1521. https://doi:10.3389/fpsyg.2014.01521

Cheung, W.-L. M. (2022). Synthesizing indirect effects in mediation models with meta-analytic methods. *Alcohol and Alcoholism*, *57*(1), 5–15. https://doi.org/10.1093/alcalc/agab044

Cichocka, A., Golec de Zavala, A., Marchlewska, M., Bilewicz, M., Jaworska, M., & Olechowski, M. (2018). Personal control decreases narcissistic but increases non-narcissistic in-group positivity. *Journal of Personality*, *86*(3), 465–480. https://doi.org/10.1111/jopy.12328

Cichocka, A., Marchlewska, M., & Biddlestone, M. (2022). Why do narcissists find conspiracy theories so appealing? *Current Opinion in Psychology*, *47*, Article 101386. https://doi.org/10.1016/j.copsyc.2022.101386

Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to "prejudice." *Journal of Personality and Social Psychology*, *88*(5), 770–789. https://doi.org/10.1037/0022-3514.88.5.770

Dyduch-Hazar, K., Mrozinski, B., & Golec de Zavala, A. (2019). Collective narcissism and in-group satisfaction predict opposite attitudes toward refugees via attribution of hostility. *Frontiers in Psychology*, *10*, Article 1901. https://doi.org/10.3389/fpsyg.2019.01901

Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual Review of Psychology*, *53*(1), 161–186. https://doi.org/10.1146/annurev.psych.53.100901.135228

Emmer, C., Dorn, J., & Mata, J. (2024). The immediate effect of discrimination on mental health: A meta-analytic review of the causal evidence. *Psychological Bulletin*, *150*(3), 215–252. https://doi.org/10.1037/bul0000419

Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests – An analysis of articles published in 2015. *Journal of Experimental Social Psychology*, *75*, 95–102. https://doi.org/10.1016/j.jesp.2017.11.008

Flores, N. M., & Huo, Y. J. (2012). "We" are not all alike: Consequences of neglecting national origin identities among Asians and Latinos. *Social Psychological and Personality Science*, *4*(2), 143–150. https://doi.org/10.1177/1948550612449025

Freis, S. D. (2018). The distinctiveness model of the narcissistic subtypes (DMNS): What binds and differentiates grandiose and vulnerable narcissism. In A. D. Herman, A. B. Brunell, & J. D. Foster (Eds.), *Handbook of trait narcissism: Key advances, research methods, and controversies* (pp. 37–46). Springer.

Fromm, E. (1947). *Man for himself: An inquiry into the psychology of ethics*. Rinehart.

Golec de Zavala, A. (2023). *The psychology of collective narcissism: Insights from social identity theory*. Routledge.

Golec de Zavala, A., Bierwiaczonek, K., & Ciesielski, P. (2022). An interpretation of meta-analytical evidence for the link between collective narcissism and conspiracy theories. *Current Opinion in Psychology*, *47*, Article 101360. https://doi.org/10.1016/j.copsyc.2022.101360

Golec de Zavala, A., Cichocka, A., Eidelson, R., & Jayawickreme, N. (2009). Collective narcissism and its social consequences. *Journal of Personality and Social Psychology*, *97*(6), 1074–1096. https://doi.org/10.1037/a0016904

Golec de Zavala, A., Cichocka, A., & Iskra-Golec, I. (2013). Collective narcissism moderates the effect of in-group image threat on intergroup hostility. *Journal of Personality and Social Psychology*, *104*(6), 1019–1039. https://doi.org/10.1037/a0032215

Golec de Zavala, A., Dyduch-Hazar, K., & Lantos, D. (2019). Collective narcissism: Political consequences of investing self-worth in the ingroup's image. *Political Psychology*, *40*(S1), 37–74. https://doi.org/10.1111/pops.12569

Golec de Zavala, A., Federico, C. M., Sedikides, C., Guerra, R., Lantos, D., Mroziński, B., Cypryańska, M., & Baran, T. (2020). Low self-esteem predicts out-group derogation via collective narcissism, but this relationship is obscured by in-group satisfaction. *Journal of Personality and Social Psychology*, *119*(3), 741–764. https://doi.org/10.1037/pspp0000260

Golec de Zavala, A., & Lantos, D. (2020). Collective narcissism and its social consequences: The bad and the ugly. *Current Directions in Psychological Science*, *29*(3), 273–278. https://doi.org/10.1177/0963721420917703

Golec de Zavala, A., Peker, M., Guerra, R., & Baran, T. (2016). Collective narcissism predicts hypersensitivity to in-group insult and direct and indirect retaliatory intergroup hostility. *European Journal of Personality*, *30*(6), 532–551. https://doi.org/10.1002/per.2067

Golec de Zavala, A., Sedikides, C., Wallrich, L., Guerra, R., & Baran, T. (2023). *Does narcissism predict prejudice?* [Unpublished manuscript]. Goldsmiths College, University of London.

Guerra, R., Bierwiaczonek, K., Ferreira, M., Golec de Zavala, A., Abakoumkin, G., Wildschut, T., & Sedikides, C. (2022). An intergroup approach to collective narcissism: Intergroup threats and hostility in four European Union countries. *Group Processes & Intergroup Relations*, *25*(2), 415–433. https://doi.org/10.1177/1368430220972178

Hase, A., Behnke, M., Mazurkiewicz, M., Wieteska, K. K., & Golec de Zavala, A. (2021). Distress and retaliatory aggression in response to witnessing intergroup exclusion are greater on higher levels of collective narcissism. *Psychophysiology*, *58*(9), 1–18. https://doi.org/10.1111/psyp.13879

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis* (2nd ed.). Guilford Press.

Jetten, J., Spears, R., & Postmes, T. (2004). Intergroup distinctiveness and differentiation: A meta-analytic integration. *Journal of Personality and Social Psychology*, *86*(6), 862–879. https://doi.org/10.1037/0022-3514.86.6.862

Leach, C. W., van Zomeren, M., Zebel, S., Vliek, M. L. W., Pennekamp, S. F., Doosje, B., Ouwerkerk, J. W., & Spears, R. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology*, *95*(1), 144–165. https://doi.org/10.1037/0022-3514.95.1.144

Leonardelli, G. J., Pickett, C. L., & Brewer, M. B. (2010). Optimal distinctiveness theory: A framework for social identity, social cognition, and intergroup relations. *Advances in Experimental Social Psychology*, *43*, 63–113. https://doi.org/10.1016/S0065-2601(10)43002-6

Lisiakiewicz, R. (2018). Poland's conception of European security and Russia. *Communist and Post-Communist Studies*, *51*(2), 113–123. https://doi.org/10.1016/j.postcomstud.2018.04.001

Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: Explaining offensive action tendencies in an intergroup context. *Journal of Personality and Social Psychology*, *79*(4), 602–616. https://doi.org/10.1037/0022-3514.79.4.602

Marchlewska, M., Cichocka, A., Jaworska, M., Golec de Zavala, A., & Bilewicz, M. (2020). Superficial ingroup love? Collective narcissism predicts ingroup image defense, outgroup prejudice, and lower ingroup loyalty. *British Journal of Social Psychology*, *59*(4), 857–875. doi.org/10.1111/bjso.12367

Marchlewska, M., Cichocka, A., Panayiotou, O., Castellanos, K., & Batayneh, J. (2018). Populism as Identity Politics: Perceived In-Group Disadvantage, Collective Narcissism, and Support for Populism. *Social Psychological and Personality Science*, *9*(2), 151–162. https://doi.org/10.1177/1948550617732393

Marchlewska, M., Górska, P., Green, R., Szczepańska, D., Rogoza, M., Molenda, Z., & Michalski, P. (2022). From Individual Anxiety to Collective Narcissism? Adult Attachment Styles and Different Types of National Commitment. *Personality and Social Psychology Bulletin*, *50*(4), 495–515. https://doi.org/10.1177/01461672221139072

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, *12*(1), 23–44. https://doi.org/10.1037/1082-989X.12.1.23

Michalski, P., Marchlewska, M., Górska, P., Rogoza, M., Molenda, Z., & Szczepańska, D. (2023). When the sun goes down: low political knowledge and high national narcissism predict climate change conspiracy beliefs. *The Journal of Social Psychology*, *164*(6), 1042–1058. https://doi.org/10.1080/00224545.2023.2237176

Obaidi, M., Anjum, G., Bierwiaczonek, K., Dovidio, J. F., Ozer, S., & Kunst, J. R. (2023). Cultural threat perceptions predict violent extremism via need for cognitive closure. *Proceedings of the*

National Academy of Sciences of the USA, *120*(20), Article e2213874120. https://doi.org/10.1073/pnas.221387412

Schmid, K., Hewstone, M., Tausch, N., Cairns, E., & Hughes, J. (2009). Antecedents and consequences of social identity complexity: Intergroup contact, distinctiveness threat, and outgroup attitudes. *Personality and Social Psychology Bulletin*, *35*(8), 1085–1098. https://doi.org/10.1177/0146167209337037

Schmitt, M. T., Branscombe, N. R., Postmes, T., & Garcia, A. (2014). The consequences of perceived discrimination for psychological well-being: A meta-analytic review. *Psychological Bulletin*, *140*(4), 921–948. https://doi.org/10.1037/a003575

Sedikides, C. (2021a). In search of Narcissus. *Trends in Cognitive Sciences*, *25*(1), 67–80. https://doi.org/10.1016/j.tics.2020.10.01

Sedikides, C. (2021b). Self-construction, self-protection, and self-enhancement: A homeostatic model of identity protection. *Psychological Inquiry*, *32*(4), 197–221. https://doi.org/10.1080/1047840X.2021.2004812

Simon, B., & Grabow, H. (2014). To be respected and to respect: The challenge of mutual respect in intergroup relations. *British Journal of Social Psychology*, *53*(1), 39–53. https://doi.org/10.1111/bjso.12019

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Vignoles, V. L., Chryssochoou, X., & Breakwell, G. M. (2000). The distinctiveness principle: Identity, meaning, and the bounds of cultural relativity. *Personality and Social Psychology Review*, *4*(4), 337–354. https://doi.org/10.1207/S15327957PSPR0404_4

Vignoles, V. L., Regalia, C., Manzi, C., Golledge, J., & Scabini, E. (2006). Beyond self-esteem: Influence of multiple motives on identity construction. *Journal of Personality and Social Psychology*, *90*(2), 308–333. https://doi.org/10.1037/0022-3514.90.2.308

Wiley, S. (2019). Perceived discrimination, categorization threat, and Dominican Americans' attitudes toward African Americans. *Cultural Diversity & Ethnic Minority Psychology*, *25*(4), 604–610. https://doi.org/10.1037/cdp0000275

Zhang, Z. (2014). Monte Carlo based statistical power analysis for mediation models: Methods and software. *Behavior Research Methods*, *46*(4), 1184–1198. https://doi.org/10.3758/s13428-013-0424-0