UNIVERSITY OF CALIFORNIA
RIVERSIDE

Seeing Myself in My Group: Generalizing From the Self-Concept to the Ingroup via Similarity and Contrastive Mechanisms.

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Jacob Johnson Elder

June 2023

Dissertation Committee:
Dr. Brent Hughes, Chairperson
Dr. Daniel Ozer
Dr. Jimmy Calanchini
Dr. Michael Erickson

The Dissertation of Jacob Johnson Elder is approved:

_____

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGMENTS

I would like to thank the members of my committee, Dan Ozer, Michael Erickson, and Jimmy Calanchini, for their time and feedback. All of them have been formative or influential in different aspects of my academic career and I am grateful that they are a part of my dissertation.

My sincere thanks go to the participants of this study, without whom this work would not have been possible. I am grateful for their willingness to participate.

In terms of my statistical knowledge and growth, I would like to thank Dan Ozer, Chandra Reynolds, and Tyler Davis. They have shaped my statistical principles, perspectives, and expertise, and I will carry their recommendations and insights into my future work. At one point in my life, I did not regard myself as a particularly quantitatively minded or qualified person, and over the past five to eight years, my expertise and interest in statistics and math have increased greatly. I owe much of this enthusiasm and growth to my statistical mentors over this period of my life.

I would like to thank the academics who have inspired me with their research. This includes Hazel Markus, Constantine Sedikides, Henri Tajfel, John Turner, Robert Nosofsky, Mark Leary and surely countless others who I forgot to list. I stand on the shoulders of academic giants who formed the seminal theories and empirical work which guide and inspire this work, and I am grateful for their influences and contributions.

I would like to thank software developers who have made my current work feasible. The improvements in computational resources in the past 10 to 20 years have been remarkable, and have allowed for the implementation of complex model fitting and

data wrangling that either would not have been possible or would have been tremendously more difficult. I owe my contributions to the people who have made these technical and computational resources so accessible and available, often without compensation. I am incredibly grateful for their contributions.

Thank you to the Psychology Department and the administration for their assisting me throughout my work on this and related projects. Thank you particularly to Sarah Turnbull, Erica Constantino, and Renee Young, for helping whenever miscellaneous issues would arise.

I am also grateful to my friends, who have provided me with support, encouragement, and inspiration throughout my academic journey. Namely, I would like to thank my cohort mates including Annie Reagan, Melissa Wilson, Laura DeLoretta, and Karynna Okabe-Miyamoto, as their continued support and understanding of shared experiences have been incredibly meaningful and have helped me throughout the process. Their friendships have made this journey more meaningful and much easier.

I would like to thank my labmates, including Genesis Morales, Eleanor Collier, Julia Hopkins, Arshiya Aggarwal, and Jennifer Mosley. Of note, I would like to highlight Bernice Cheung who contributed much of the early work to develop and implement the trait network and was the driving force in implementing a fMRI preprocessing pipeline at the start of my PhD, and with setting up much of the lab infrastructure more generally. These tasks and the processes she implemented early on have been foundational to my work and I could not have done it without her. I would also like to thank Yrian Derreumaux—As a friend and labmate who was always willing to work through

problems at the drop of a hat, provide attention to personal or work stressors, and be available more generally. Without Yrian, I question whether I would be here now at the end of my PhD.

I would like to express my gratitude to my family, who have always been there for me, offering love, support, and encouragement throughout my academic pursuits and throughout my life more generally. Their unwavering faith in me has been a constant source of motivation and inspiration. Thank you Luke Elder, Anne Elder, and Brent Elder, for all that you have done and continue to do for me.

I would like to express gratitude to my supervisor, Brent Hughes, for his invaluable guidance, support, and encouragement throughout the course of this project and my PhD. His insights and feedback were instrumental in shaping my research and helping me stay on track. He has continually provided advice on my career, and provided prompt feedback on even the smallest professional items such as inquiry emails or poster submissions. He has also continued to prioritize weekly meetings with all of his graduate students. Together, doing this for many graduate students at once I imagine is quite time consuming and demanding. I have appreciated his dedication to his students and to being personally and professionally supportive of his students. I have enjoyed our weekly meetings and are continued discussions about research and life. I have also appreciated the community and culture of our lab that he has developed. I appreciate him for being understanding, helpful, and supportive, and ultimately for taking a shot on me as a PhD student. I would not be at this stage without him having taken a gamble on me and for believing in me.

I would like to also express gratitude for my second mentor and pseudo-advisor during my PhD, Tyler Davis, who has been involved with every project of my PhD. Nearly all of my interactions with him have been remote, but he has been immensely influential and instrumental in my learning about computational modeling, cognitive science, statistics, and fMRI methods. Much of my current perspectives on methods and statistics I owe to Tyler and his insights. I am grateful for his long-winded and well-thought-out emails to my also long-winded emails, which are responsible for much of my technical growth as a researcher and professional. He has managed to continue to work on our papers and respond to my emails even despite maintain a full-time job outside of academia, which I remain astounded and impressed by. Despite often disagreeing with Tyler on various details, he would nevertheless engage with my disagreements and push me. I would not have formed the research identity that I currently have without Tyler's assistance and education throughout the process, nor would I have the confidence in my knowledge without having to so frequently make my case or defend my perspective. For this opportunity to learn and grow, I am grateful.

Thank you all for your support, encouragement, and guidance throughout this journey. Your contributions have been invaluable, and I am truly grateful for everything you have done for me.

ABSTRACT OF THE DISSERTATION

Seeing Myself in My Group: Generalizing From the Self-Concept to the Ingroup via Similarity and Contrastive Mechanisms.

by

Jacob Johnson Elder

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, June 2023
Dr. Brent Hughes, Chairperson

People tend to see themselves as similar to their ingroup, and people often accomplish similarity with others by projecting their self-beliefs onto their perceptions of others. However, existing research on *self-anchoring* has not considered the within-person cognitive mechanisms facilitating this process. The current study aims to establish the similarity-based (i.e., if I am outgoing, my group ought to be characteristic of semantically similar traits such as sociable and fun) and contrastive (i.e., what is characteristic of my ingroup in contrast to a given outgroup) mechanisms by which people's self-evaluations on traits generalize to ingroup evaluations. Across three studies using minimal groups (N = 61), university groups (N = 283), and racial groups (N = 265), we find that people use semantic similarity among traits to infer the extent to which traits

ought to be characteristic of their group if related traits are characteristic of themselves. We further find that this tendency is primarily driven by a motivation to achieve similarity with the ingroup rather than dissimilarity from the outgroup. However, in the racial context, racial minority participants contrasting against the racial majority were driven moreso to achieve dissimilarity from the majority outgroup. We fit a computational model measuring the extent to which people convert self-beliefs into ingroup-beliefs prior to generalization, and find that this tendency was weaker when people contrasted their ingroup against an outgroup that they felt more positively about (i.e., the higher status university in Study 2 and the fellow minority racial group is Study 3), reflecting that self-anchoring may be more pronounced when contrasting against majority or more disliked outgroups. In fact, this projection rate was correlated with self-reported intergroup bias in studies 2 and 3 and social identification in all three studies, reflecting that the extent to which individuals generalize about their groups based on themselves may depend on how biased and affectively attached they are to their social groups. Findings reflect that how people generalize from the self to the group may enact similarity-based classification processes that are amplified under particular intergroup contexts.

Table of Contents

List of Figures

Seeing Myself in My Group: Generalizing from the self-concept to the ingroup via similarity and contrastive mechanisms

People carry their distinct sets of stable self-beliefs along with them throughout various situations (Markus & Wurf, 1987), yet also manage to achieve a sense of belonging (Baumeister & Leary, 1995) by assimilating to their various social groups (Brewer, 1991; Ellemers et al., 2001). In achieving a sense of similarity with one's social groups, people promote a sense of positive attachment, or social identification, with their groups (Tajfel, 1978; Turner et al., 1987). People tend to represent these group memberships (i.e., ingroups) as compatible with their self-concepts (Smith & Henry, 1996), and greater perceived overlap between oneself and one's ingroup tends to beget greater social identification (M. Cadinu & De Amicis, 1999; Coats et al., 2000; Tropp & Wright, 2001). While one route to achieving this social identification might be by assimilating group attributes into the self-concept (Turner et al., 1987), this dominant account for group assimilation does not sufficiently explain how people identify with novel groups or how people maintain self-beliefs that are stable beyond group memberships (van Veelen, Otten, et al., 2016). Rather than merely limitlessly assimilating attributes of one's ingroup into the self-concept, people also project their own attributes onto how they represent and perceive their ingroups (M. Cadinu & Rothbart, 1996) as well as others in interpersonal contexts (Ames, 2004). However, the within-person mechanisms by which individuals self-project– or generalize their own self-beliefs– to their ingroup are not yet fully clear. Specifically, people may consider the similarity relations among self-beliefs when inferring how they ought to characterize and

generalize to their group, as well as contextual factors, such as what group people compare their ingroup against, may augment how people generalize.

Inferences About the Ingroup Based on the Self

The self serves as an informational base (Gramzow et al., 2001) from which people draw inferences about their ingroup, otherwise known as self-anchoring. For example, there is a strong association between the positively represented self-concept and one's ingroup (Clement & Krueger, 2002; DiDonato et al., 2011; Gramzow & Gaertner, 2005) and self-evaluations account for group evaluations to a greater extent than the mere social desirability of traits (Clement & Krueger, 2000, 2002; Otten & Wentura, 2001). Specifically, individuals engage in inductive reasoning and infer unknown information about their group on the basis of their own self-knowledge (DiDonato et al., 2011; Krueger, 2007), generalizing from themselves to a multitude of group members. As such, self-anchoring is most likely to occur in situations in which group knowledge is unknown or unclear (van Veelen et al., 2013a), such as minimal groups without acquired or diagnostic group knowledge.

However, despite the apparent strength of evidence for people generalizing to group evaluations from self-evaluations, the majority of the research in this domain has relied on trait ratings for the self followed by trait ratings for the group, and the concordance of each pair of trait ratings across the self and ingroup is compared using distance (M. Cadinu et al., 2020; M. Cadinu & Rothbart, 1996) or correlational (Bianchi et al., 2009; Otten & Wentura, 2001; Sherman & Kim, 2005; van Veelen et al., 2011) measures. Thus, these prior tests of self-anchoring provide evidence that people evaluate

similarly on the same traits across the self and group, but not that people evaluate similarly on similar but different traits across the self and group. In addition, it is the case that the mere repetition of information causes it to be perceived as more true or characteristic (Unkelbach, 2007; Unkelbach et al., 2019; Unkelbach & Rom, 2017), and one critique of this prior work may be that it relies strictly on the correlations among repeated ratings. As such, this prior work relying on correlations or distances among repeated ratings may provide a test of stability in ratings, but not necessarily generalization per se. A stronger test of this generalization-based theory of self-anchoring and -projection would be to establish that this generalization occurs across traits, to novel traits that are not merely repeated observations. As such, here we attempt to establish that people's self-evaluations on traits can generalize to group evaluations on novel but related (Heit & Rubinstein, 1994) traits.

Relational Similarity as the Basis for Self-Concept Generalization

The stimuli, people, and situations encountered by an individual are likely to vary considerably across experiences, which necessitates that people be adept at generalizing to new stimuli, persons, and situations on the basis of similarity to prior experiences (Shepard, 1987). By extension, when evaluating ingroups, people may generalize that unobserved ingroup members may be like them, based on the belief that group members are bound together by similar attributes (R. J. Brown, 1984). However, beyond merely inferring that one is similar to one's group on the same traits (i.e., if I am outgoing, my group is also outgoing), people may also infer similarity with one's group on similar traits which they have not yet or recently self-evaluated on (i.e., If I am outgoing, my

group ought to be sociable, funny, and fun). In recent research, we have developed a semantic network of trait dependencies (J. Elder, Cheung, et al., 2023), that allows for the extraction of relational similarity among pairs of traits based on common neighbors in the network. This network model of trait relations that contains information about relational similarity is thus useful for identifying people's representations of themselves and their social groups, and how they generalize on the basis of similarity among trait relations.

The usefulness and robustness of network-derived semantic similarity is well-established, as we have used these similarity relations in prior work to examine how the brain represents semantically similar traits during self-evaluations (J. Elder, Cheung, et al., 2023), how feedback propagates across traits as a function of similarity (J. Elder, Davis, et al., 2023b; J. Elder et al., 2022c), how similar self-evaluations among similar traits predicts confidence in self-evaluations (J. Elder et al., 2022a), how people assimilate group norms into the self-concept (J. Elder et al., 2022b), and how people reflect on themselves as similar to others (Schneider et al., 2022). Relevantly, inferred similarity is not only important for classifying elements of a particular category together, but also facilitates the contrasting of distinctive features among stimuli (Tversky, 1977). Specifically, the extent to which social groups are represented as different from one another and compatible with oneself may depend on situational factors, such as how they are contrasted against one another.

Contrastive Principles Augment Category Representations

People generally accentuate differences between their ingroup and relevant outgroups (Tajfel et al., 1964; Tajfel & Billig, 1974) and similarities within social groups

(Haslam et al., 1995; Turner et al., 1994). Such differences between groups and similarities within groups may be magnified under conditions in which one's ingroup and other outgroups are contrasted against one another. This premise was formalized in early social identity research using the metacontrast principle, which defined the likelihood of a given individual being categorized as a group member (i.e., the ostensible self-prototypicality of one's ingroup) as the ratio of the individual's similarity to the ingroup relative to the individual's similarity to the outgroup (Turner et al., 1987). Indeed, for non-social group related categories, contrasting opposing categories against each other causes their mental representations to be repelled and the resulting estimates and beliefs about each category to be polarized (Davis & Love, 2010; Vogel et al., 2018), and contextual factors can alter how the similarity relations among stimuli are represented (Nosofsky, 2011). However, while this metacontrast principle has been verbally described in terms of similarity, little to no research has formally implemented this principle in the context of social categories using relational similarity measures (Davis & Goldwater, 2021).

More generally, despite an abundance of representational and mechanistic claims in the intergroup and intragroup processes literature that are rooted in cognitive science theory on category and concept representations, including the claim that people consider themselves as interchangeable exemplars of their social group prototype (Hogg et al., 1995, 2004), little research has implemented formal category learning models to test these principles. Given the network approach implemented here, we are able to provide some of the first formal tests of intragroup theory rooted in category learning models, in a

self-anchoring context. Prior research has established that intergroup salience promotes tendencies to self-anchor (Krueger & Clement, 1996), and we expect that ingroups compared against outgroups will accentuate intergroup salience and enhance the tendency to self-anchor.

The Current Design

Concept generalization is commonly established by having participants learn the features of concepts based on a subset of concept exemplars (i.e., examples), and then generalizing in a test phase to a different set of concept exemplars (Bowman et al., 2020). For example, cartoon animals may differ along multiple binary dimensions such as color (yellow/gray), shape (squared/circular), and orientation of dots on body (vertical/horizontal), and concept examples that share the most features with the prototypical animal 'A' are most likely to be classified as animal 'A'. Consistent with this framework, the current study involves a "training phase" whereby participants self-evaluate on a subset of traits which can be thought of as exemplars of the self-concept prototype. This is then followed by a "generalization phase" whereby participants classify a trait as more characteristic of the ingroup or outgroup. Using the traits' semantic similarities to the self-concept, we can estimate whether they will be classified as belonging to the ingroup or outgroup category. In doing so, we are able to predict the likelihood of a trait being characterized as typical of one's ingroup, both for traits that were previously self-evaluated during training (i.e., *repeated* traits) and for traits that were not self-evaluated during training (i.e., *novel* traits). We predict that participants will generalize to the ingroup, and infer not only that the ingroup is similar to them on traits

they evaluated on, but on novel traits that are semantically similar to the traits they evaluated on. Additionally, we test different intergroup contexts which may amplify or attenuate the self-descriptiveness weights that people use for similarity-based generalization.

Study 1: Self-Anchoring Based on Minimal Conditions

In contexts in which the ingroup is not clearly defined, people infer characteristics of the ingroup on the basis of the self (M. Cadinu & Rothbart, 1996; van Veelen, Otten, et al., 2016). Minimal groups are a well-established paradigm for promoting group identification and ingroup favoritism, even under conditions in which knowledge of the group is "minimal" beyond one's knowledge of belonging to the group (Otten, 2004). We first sought to test the similarity and intergroup contrast mechanisms in a group context in which self-anchoring effects should be most likely to occur, specifically by testing whether people will self-project onto a minimal ingroup which people have no prior diagnostic knowledge about, such as a randomly assigned "underestimator" or "overestimator" ingroup label. This should thus provide a first proof-of-concept test that people can generalize to the ingroup using semantic similarity.

Methods

Participants

We recruited 80 participants and excluded 19 participants to arrive at a final sample of N = 61. We excluded any participants who indicated that they did not consent to their data being used after the debriefing (N = 9), who self-reported not taking the task seriously (less than '4' to "To what extent did you take this task seriously?"; N = 1), who

self-reported that their data is unusable ("No" to "Did you understand the task and respond truthfully and meaningfully enough that your data is usable?"; N = 2). We additionally excluded participants if they exhibited behavior during the task that reflected careless responding, with exclusionary criteria including if over 80% of their self-evaluations were identical, if over 95% of their group classification were identical, or if over 40% of their behavioral responses were missing from either part of the task (N = 10).

Participants (N = 61) were native English-speaking university students (40.98% Cisgender Female, 59.02% Cisgender Male; M_{Age} = 19.95, SD_{Age} = 1.51, $Range_{Age}$ = [18, 25]), and were 9.84% White/Caucasian, 11.48% Mixed/Other, 32.79% Asian, 40.98% Hispanic/Latino, and 4.92% Native-American.

We conducted a power curve analysis on a prior dataset involving relational similarity, group processes, and traits (J. Elder, Cheung, et al., 2023) using *simr* (Green & MacLeod, 2016), predicting self-evaluations after learning from feedback-weighted similarity to prior traits. The power curve analysis revealed that a sample size as small as N = 5 was sufficient to detect the within-subjects effect at 95% power, but we aimed to recruit a larger sample. The primary focus of this study was the within-subjects inferences which require smaller sample sizes to be sufficiently powered.

Network Procedure

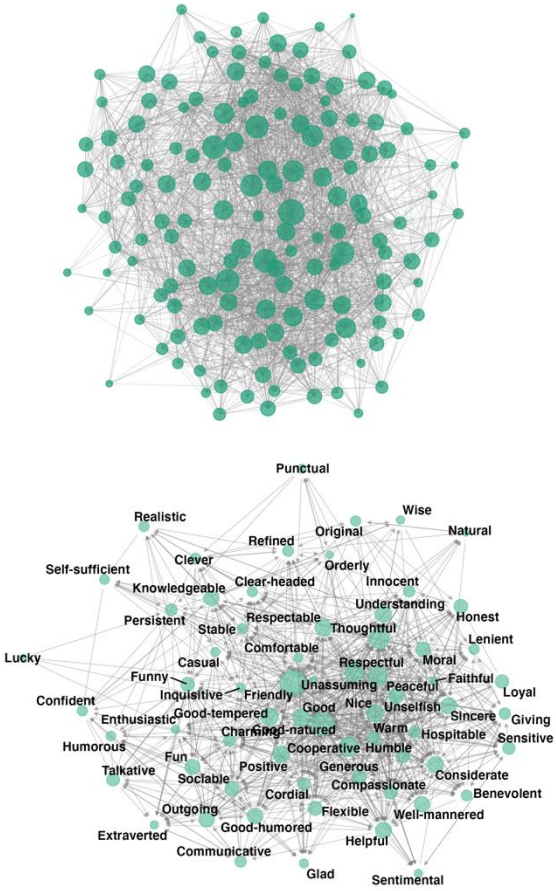
178 Amazon Mechanical Turk participants contributed to the construction of the positive trait network. Each participant nominated which of 147 other traits depended

upon a target trait for semantic meaning. If there was consensus among a sufficient number of participants (i.e., 25%), the dependency relation was included as a connection in the network. From this we generated a directed dependency network (Figure 1) of trait relations (Elder et al., 2023 for more detail). From this network, we derived network-based Dice similarity:

$$Dice = \frac{2*|A \cap B|}{|A|+|B|}, [1]$$

where A and B denote any two given traits in the network, the numerator denotes twice the number of their common neighbors, while the denominator denotes the sum of their total connections. More colloquially, this similarity measure represents the proportion of overlap between two traits based on their number of shared features (i.e., neighbors), consistent with feature-based models of semantics (Martin, 2007; Tyler & Moss, 2001) and similarity (Tversky, 1977). Additionally, we derived outdegree centrality, which was defined as the number of traits that depend on a given trait (sum of a given trait's row in the adjacency matrix; how many of columns j depend on row i). We also derived indegree centrality, which was defined as the number of traits a given trait depends on (sum of a given trait's column in the adjacency matrix; how many of rows i column j depends on).

Figure 1. Trait network.



Note. Top figure depicts full semantic dependency network of traits. Bottom figure is subset of network containing the trait 'Friendly' and it's immediate neighbors.

Design

Training phase. Participants were provided consent and subsequently completed demographics questionnaires. self-evaluated on approximately 60% (90 traits) of all 148 traits within the trait network of semantic dependency relations (J. Elder, Cheung, et al., 2023). At each trial, participants evaluated the extent to which each trait is self-descriptive on a Likert scale from 1 (Not at all) to 7 (Extremely), and trials terminated upon participant response. This phase of the task is considered the "training" phase.

Minimal group assignment. After self-evaluating on each trait, participants underwent a minimal group assignment (Hong & Ratner, 2021; Otten, 2004). Specifically, participants were told:

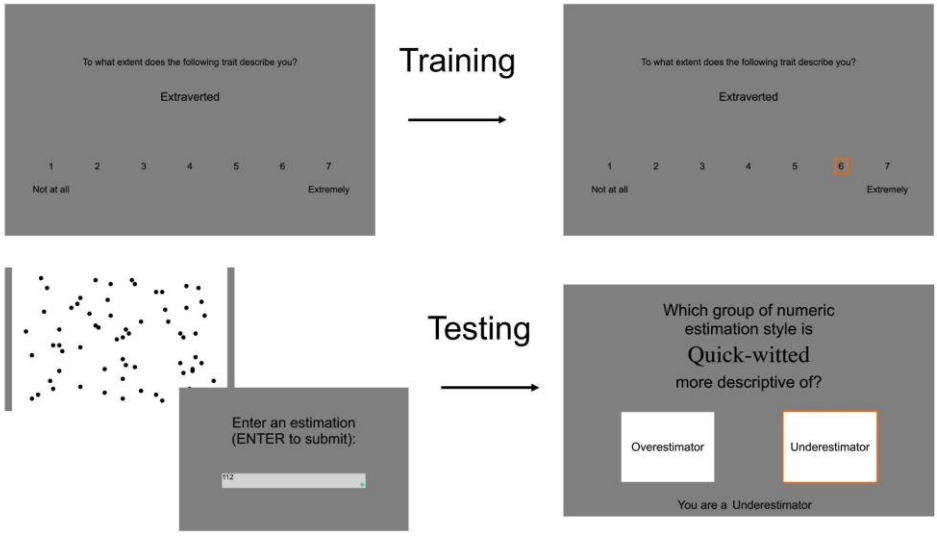
> *"People vary in numerical estimation style, or the tendency to overestimate or underestimate the number of objects one encounters. Approximately half the population are overestimators and half are underestimators, and there is no relationship between numerical estimation style and any other cognitive tendencies. We will ask you to complete a well-established task called the Numerical Estimation Style Test (NEST) to determine what type of numerical estimation style you are. We will then ask you to evaluate a multitude of traits and determine which type of numerical estimation style each is more likely to be characteristic of."*

The participants estimated the number of dots on screen for 10 trials of dots appearing for 300ms. Subsequently, participants were assigned to be either overestimators or underestimators via counterbalancing. In order to amplify social identification with the assigned minimal ingroup, participants were asked to confirm their ingroup by typing as a response, "I am a [ESTIMATION STYLE]."

Generalization phase. Participants then underwent a two-alternative forced choice task in which they observe each trait from the network, and choose whether it is more descriptive of overestimators or underestimators (i.e., the ingroup or the outgroup, depending on their random assignment), across 148 trials. The trait was presented on the screen, "Which group is [TRAIT] more descriptive of?". At the bottom of the screen, a constant reminder is provided of the minimal ingroup, "You are a [ESTIMATION STYLE]". This phase of the task is considered the "generalization" phase (see Figure 2 for task schematic). There were 58 traits that were not observed during the training phase

that were then observed during the generalization phase, providing a set of "novel" traits which were not previously self-evaluated and thus could only be classified on the basis of their similarity to other traits (Figure 2).

Figure 2. Schematic of task design



Note. Depicting training phase on top and generalization phase on bottom. Participants self-evaluate on 90 out of 148 traits on a Likert scale from 1 to 7. During generalization phase, 90 traits are thus repeated while 58 traits are novel and never received a self-evaluation. Participants then proceed through all traits in the semantic trait network, and classify whether each trait is characteristic of the ingroup or the outgroup.

Indices

We computed a variety of metrics using our relational similarity measures, in order to understand how people generalize from the self to the ingroup.

Similarity-to-Self. In order to measure the extent to which a trait is similar to prior self-evaluations, we first computed a *Similarity-to-self* measure, which incorporates information about both the self-descriptiveness and similarity of traits observed during

the training phase. A trait's Similarity-to-Self, SS_g, is determined by its similarity to prior traits and their self-descriptiveness:

$$SS_g = \frac{\sum_{t=1}^{T=90} E_t * Sim_{tg}}{\sum_{t=1}^{T=90} E_t}, [2]$$

This is a self-evaluation-weighted average of a given trait's similarity to the training phase traits. S_{tg} is the similarity of the group-evaluated trait *g* at the generalization phase to the self-evaluated trait *t* at the training phase (subscript *tg* denotes a pairwise relationship between trait *t* and *g*), and E_t is the self-descriptiveness of each trait from the training phase. This measure attenuates or amplifies the similarity of a given trait to prior traits as a function of its self-descriptiveness, such that if outgoing is being classified during generalization, and similar traits such as sociable, fun, and witty were evaluated as self-descriptive previously, the similarity-to-self will be higher while if they were evaluated as less descriptive, the similarity-to-self for outgoing will be lower. Given this measure will be higher for traits that are similar to self-descriptive traits and dissimilar from non-descriptive traits, it should thus be predictive of a trait's likelihood of being classified as ingroup characteristic.

Uncertainty. Uncertainty is theorized to be an integral motivator underlying group identification (Hogg, 2007, 2014), such that social identification is thought to be most likely in contexts when people are most uncertain of themselves. However, little to no research has formally or mathematically defined uncertainty (Shannon, 1948) in the context of group identification, to more precisely measure the extent to which uncertainty

relates to group identification. Given a set of probabilities reflecting the likelihood of different self-evaluation responses, we can estimate a measure of overall uncertainty using a standard entropy formulation (Davis et al., 2012a, 2012b; J. Elder, Davis, et al., 2023a). Uncertainty represents the likelihood of any self-evaluation response from 1 to 7 given prior traits evaluated, such that more uncertainty may be represented by equivalent likelihoods across all feedback categories as follows:

$$Entropy = -\sum_{E}^{K} P_{E\,g} * log_2 P_{E\,g}, [3]$$

where $P_{E\,g}$ denotes the probability of self-evaluating E for trait g, and E is one of K self-evaluation response categories possible. Here, the probability of evaluating one of K possible responses is computed as the summed similarity of the current trait to all prior traits that received that feedback over the summed similarity of all prior traits regardless of evaluations. Thus, the probability that trait g is evaluated as E is defined as follows:

$$P_{E\,g} = \frac{\sum_{e \in E} S_{ge}}{\sum_{K} \sum_{k \in K} S_{gk}}, [4]$$

where S_{ge} represents the similarity of the trait g (observed during generalization) to trait e that received self-evaluation E, and the index $e \in E$ indicates that the sum is over all traits e that were rated E. The denominator sums over all responses categories K. In this uncertainty formula (FeldmanHall & Shenhav, 2019; Hirsh et al., 2012), if all self-evaluation responses were equally likely because of all prior self-evaluations being of equivalent similarity to the current trait, the current trait group classification, g, would

have higher uncertainty. Conversely, if the current trait is most similar to traits that received self-evaluations '6' and '7', but not similar to traits that received other types of self-evaluations, the current trait *g* would have lower uncertainty.

Similarity-to-Group and Metacontrast Ratio. Unlike the previously described indices, we additionally computed indices for traits self-evaluated on during the learning phase. Each trait self-evaluated on was later either predicted to belong to the outgroup or the ingroup. Therefore, for each trait self-evaluated on during the learning phase, we estimated the summed similarity to all traits predicted as belonging to the ingroup or the ougroup within each participant:

$$SG_t = \sum_{i \in I} S_{ti}, [5]$$

where SG_t is the similarity of ingroup-classified trait *i* (out of all ingroup classified traits I) to the self-evaluated trait *t*. This is estimated separately for the ingroup (Similarity-to-Ingroup) and outgroup (Similarity–to-Ingroup).

The metacontrast principle (Turner et al., 1987), suggests that an individual's prototypicality of the ingroup and likelihood of assimilation can be defined by the ratio of the individual's average similarity to the ingroup over the individual's average similarity to the outgroup. This premise has been verbally expressed but rarely or never formally tested using relational similarity measures, which we implement here:

$$MCR_t = \frac{\sum_{i \in I} S_{ti}}{\sum_{o \in O} S_{to}}, [6]$$

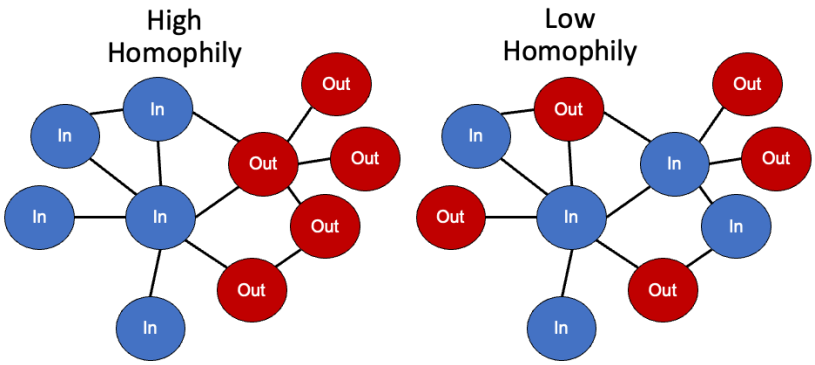
where S_{ti} denotes the similarity of trait *t* evaluated during training to trait *i* out of all traits *I* that were classified as characteristic of the ingroup during generalization (i∈I), and Sim_{to} denote the similarity of trait *t* evaluated during training to trait *o* out of all traits

O classified as characteristic of the outgroup during generalization ($o \in O$). While the original metacontrast ratio in Self-Categorization Theory was conceptualized as using "average" similarity, in the concept and category learning literature, it is more conventional to use summed similarities for applications such as this, as sums encode frequency information. Sums better reflect that people store particular instances in memory, which are accumulated to translate to choice likelihoods (Don et al., 2019; Don & Worthy, 2022; Estes, 1976a, 1976b). This formula thus denotes trait *t's* (observed during training) summed similarity to the ingroup relative to the outgroup. Thus, to the extent that the metacontrast ratio is higher, this reflects that a trait is more similar to the ingroup and/or more dissimilar from the outgroup. Traits with a higher metacontrast ratio should be evaluated more self-descriptively in general, as it reflects that traits that were more self-descriptive were more similar to ingroup classifications and more dissimilar from outgroup classifications.

Trait Segregation by Group. While all the other network indices were trait-level measures, we additionally computed a measure reflecting the extent to which individuals separate traits that are classified as characteristic of the ingroup or outgroup in their mental representations of these groups, which we label as Trait Segregation. Specifically, for all traits in the network, participants classified whether they belong to the ingroup or the outgroup. Thus, each trait within the network has a categorical label of outgroup or ingroup assigned to it. We then compute a measure of network nominal homophily, reflecting whether ingroup/outgroup traits tend to connect with other ingroup/outgroup traits. Higher/more positive values reflect that connected traits tend to have the same

group label, while lower/more negative values reflect that connected traits tend to have different group labels (Figure 3).

Figure 3. Depiction of trait segregation.



Note. A more homophilous participant is someone who assigned highly interconnected traits an ingroup classification and separate interconnected traits an outgroup classification. A less homophilous participant is someone who intermixed ingroup and outgroup classifications among the network of interconnected traits.

Generalization Model

The previously described *Similarity-to-self* measure can provide an indication of whether a trait's semantic similarity to prior trait self-evaluations may be associated with its likelihood of ingroup classification. However, they may not provide insight into the underlying psychological processes and mechanisms. To further elucidate these mechanisms, we implement a computational model emulating classic concept learning and generalization models (Maddox & Ashby, 1993; Nosofsky, 1984, 1988, 2011).

Bias Model. As a first model, primarily to be used as a reference or comparison, we implemented a model whereby individuals differ in their tendency to be biased towards ingroup or outgroup choices on average.

$$P\,(Ingroup \mid g) = \frac{\gamma^{\beta}}{\gamma^{\beta} + (1-\gamma)^{\beta}}, [7]$$

where the bias parameter, γ, reflects an overall tendency to classify traits as ingroup-typical (more towards 1) or outgroup-typical (more towards 0). γ was allowed to vary from 0 to 1. This model uses 1 free parameter and does not provide insight into the mechanisms by which people use self-beliefs to project to the ingroup or reject the outgroup.

Self-Projection Model. As a second model, we tested the extent to which people self-anchor, or project their self-evaluations onto the ingroup relative to the outgroup. To do so, we implemented a logistic function which transforms self-evaluations (E_t) into ingroup-beliefs ($InGB_j$) on training trial *t*:

$$InGB_t = \frac{1}{1 + e^{-\alpha*(E_t-4)}}, [8]$$

where E_t denotes participant self-evaluations from the training phase of the task, α indicates a participant's *projection rate*, reflecting the extremity with which self-beliefs are converted to ingroup-beliefs. α was allowed to vary from 0 to 10. Self-evaluations are centered at 4, which is the midpoint on the scale. A higher α reflects more extremity (i.e., more sigmoidal) in the tendency to project self-to-ingroup, such that a rating of '5' is perceived as highly ingroup characteristic, whereas a lower α reflects less extremity (i.e., more linear), such that the transformation from self-evaluations to ingroup beliefs is relatively equivalent.

Outgroup beliefs ($OutGB_t$) are assumed to be the opposite of ingroup beliefs, such that the *projection rate*, α, is flipped for the outgroup:

$$OutGB_t = \frac{1}{1 + e^{\alpha*(E_t-4)}}, [9]$$

As such, the extent to which the projection rate predicts more ingroup projection by extension also predicts corresponding outgroup rejection. Ideally, we would have liked to fit separate parameter for the ingroup and outgroup projection rates, but the parameters were not identifiable when attempting to fit this. This is likely due to an insufficient amount of available data from the training phase to distinguish ingroup and outgroup from a single set of self-evaluations. The best solution turned out to be to allow the projection rate for ingroup and outgroup beliefs to be in opposing directions.

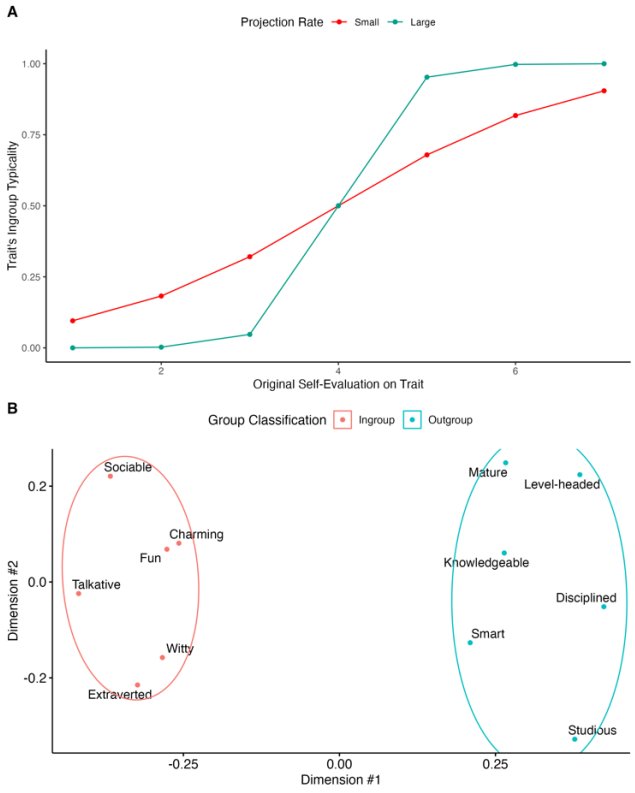
The probability of ingroup classification during generalization trial *g* is depicted by:

$$P(Ingroup \mid g) = \frac{\gamma * \sum_{t=1}^{T}(S_{tg} * InGB_t)^{\beta}}{(1-\gamma) * \sum_{t=1}^{T}(S_{tg} * OutGB_t)^{\beta} + \gamma * \sum_{t=1}^{T}(S_{tg} * InGB_t)^{\beta}}, [10]$$

where S_{tg} denotes the similarity between the trait *g* observed on the current generalization trial and training trait *t* observed on a given training trial, β denotes the *temperature* parameter which governs the stochasticity (lower values) or determinism (higher values) with which group classifications are made, and γ denotes bias which amplifies the likelihood of a trait being classified for the ingroup (higher) or outgroup (lower). β was allowed to vary from 0 to 10. The ingroup beliefs and outgroup beliefs for each trait observed on training from *t* through *T* are multiplied by their similarity to the current trait observed on generalization trial *g* and summed to denote a measure of ingroup-typicality and outgroup-typicality respectively (Nosofsky, 1988, 1991). The higher the ingroup-typicality (given ingroup projection and outgroup rejection of self-

beliefs), the greater the likelihood of ingroup-typicality. A depiction of how the *projection rate* changes across different values of α is depicted by Figure 4.

Figure 4. Depiction of projection rate and group classifications of semantically similar traits.



Note. (A) Visualization of how *projection/rejection rate* from self-to-group varies across different values of α (Small = .75, Large = 3.0). (B) The network similarity relations among 12 traits, depicted using multidimensional scaling (MDS). K-means clustering is performed to classify traits into clusters. As an example, the k-means clusters are labeled "ingroup" and "outgroup" to denote how an example participant may infer that different types of traits are more characteristic of the ingroup or the outgroup. MDS is a method for depicting the similarity relations among elements in a dataset. MDS is most commonly performed to extracted two dimensions so that similarity relations can be intuitively visualized in two-dimensional-space. The x and y-axis are thus the two dimensions extracted from MDS based on the similarity relations.

Model fitting. For model fitting, we used the probabilistic programming language Stan, which uses Markov chain Monte Carlo (MCMC) sampling algorithms. Hierarchical

Bayesian analysis (HBA) (Huys et al., 2011) was implemented by emulating the procedure detailed by *hBayesDM* (Ahn et al., 2017), which stabilizes and regularizes individual-level parameter using group-level estimates (Ahn et al., 2011). All of the models were fitted for each subject in the study. Posterior parameter distributions were sampled for each subject. A total of 1000 samples were drawn after 1000 burn-in samples (overall 2000 samples) in four MCMC chains. We assessed if MCMC chains converged to the target distributions by inspecting *R*hat values for all model parameters, and checking if Rhat values were less than 1.01 (more strict) or 1.05 (more liberal) (Vehtari et al., 2021). Posterior distributions for all parameters for each of the subjects were summarized by their median as the central tendency resulting in a single parameter value per subject that we used to calculate group statistics.

Model selection. In order to evaluate the winning model, we estimated pointwise out-of-sample prediction accuracy for all fitted models separately for each participant by approximating leave-one-out cross-validation (LOO) as recommended for assessing model fit. Specifically, we used the Bayesian LOO estimate of the expected log pointwise predictive density (Vehtari et al., 2017).

Individual Differences Measures

We collected participants' self-report on a variety of measures of individual differences measures, including dialectical self-views (Spencer-Rodgers et al., 2015), self-esteem (Rosenberg, 1965), self-concept clarity (J. D. Campbell et al., 1996), need for

cognition (Cacioppo & Petty, 1982), self-prototypicality of ingroup (single-item; (Fielding & Hogg, 1997; Hogg & Hains, 1996), need to belong (Leary et al., 2013), and social identification (single-item) (Postmes et al., 2013).

Planned Analyses

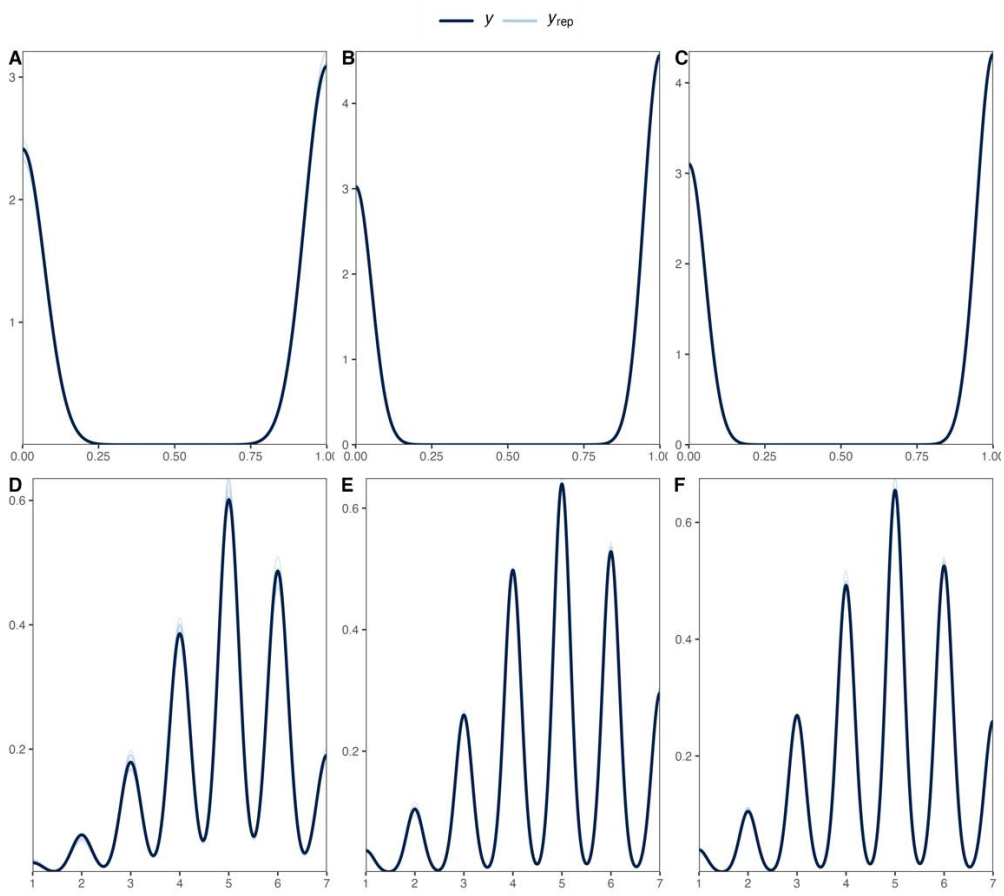
For analysis and cleaning we used R Programming Environment 4.2.1. Bayesian generalized mixed-effects models were estimated in *brms* using Cumulative Link Mixed Models (CLMM) for Likert outcomes and binomial generalized linear mixed models (GLMMs). The logit link function for the CLMMs and GLMMs generate log odds estimates, which we additionally exponentiate and report as Odds Ratios (ORs) for ease of interpretation. For all models, we implemented crossed random factors, with a random factor modeled for both subjects and traits (Baayen et al., 2008). Traits were modeled as random factors in addition to subjects, as omission of stimulus-level variation in random effects specification can lead to substantially inflated false positive rates, with psychological experiments using only by-participants random effects averaging a Type I error rate of 23.90% (Judd et al., 2012; Yarkoni, 2022). In describing random effects estimated for intercepts and slopes, we describe the intercepts and slopes as "allowed to vary" or "estimated as varying" for more intuitive language (Bafumi & Gelman, 2007).

Fully Bayesian models can have advantages over Frequentist models estimated with maximum likelihood particularly for more complex models with outcomes distributed as an ordinal, beta distribution, or a mixture of distributions. Additionally, it is

generally recommended to estimate maximal random effects and to remove as needed to control the Type I error rate (Barr et al., 2013), but when using maximum likelihood Frequentist approaches, fitting the maximal random effects structure often results in non-convergence or aberrant random effects (e.g., perfect correlations among random effects). In contrast, the maximal random effects structure can be readily estimated in a Bayesian framework (Eager & Roy, 2017; Nicenboim & Vasishth, 2016; Sorensen et al., 2016). Further, the Bayesian framework comes with other advantages, for instance, the ability to derive probability statements for every quantity of interest or explicitly incorporating prior knowledge about parameters into the model. Following the Sequential Effect eXistence and sIgnificance Testing framework (Makowski et al., 2019), we report the median of the posterior distribution and its 95% CI (Highest Density Interval), and the probability of direction (*pd*) which quantifies the certainty with which the effect is positive or negative. This is akin to a frequentist p-value, which we report for convenience the approximation: 2 * [1 - *pd*]. We also report the probability of significance (*ps*; $\beta >$ | .05 |), which reflects the probability that effect is above a given threshold corresponding to a negligible effect in the median's direction. Finally, we report the probability of the effect being large (*pl*; $\beta >$ | .30 |). We use the *brms* default for priors with flat, uninformative priors for fixed effects, and weak Student's T distributed priors for the random effects in order to weakly regularize the posterior without biasing the effect by prior.

We conducted posterior predictive checks (PPCs) are conducted to compare the predicted and observed data, and to support that the models are well-suited to describe the data. To the extent that data simulated from the posterior predictive distribution (Y_{rep}) resembles the observed outcome data (Y). It is useful in evaluating model adequacy and whether the predictions are valid. We depict several PPCs from the training and generalization phase for each study to support that our models meet assumptions and are appropriate for the observed outcome data (Figure 5). As is apparent from the figures, the observed data and predicted data are well-aligned, demonstrating model-adequacy for generating predictions.

Figure 5. Posterior predictive checks for primary models in each study.



Note. Y is predicted data and Y_{rep} is simulated data. (A) Ingroup classifications predicted by Similarity-to-self in Study 1. (B) Ingroup classifications predicted by Similarity-to-self and Condition in Study 2. (C) Ingroup classifications predicted by Similarity-to-self and Condition in Study 3. (D) Likert self-evaluations predicted by Metacontrast Ratio in Study 1. (E) Likert self-evaluations predicted by Metacontrast Ratio and Condition in Study 2. (F) Likert self-evaluations predicted by Metacontrast Ratio and Condition in Study 3.

Effect of Desirability on Group Predictions. We first investigated whether the social desirability[1] of traits predicts whether people will choose a trait as characteristic of the ingroup or not. We regressed ingroup classification (Ingroup = 1/Success; Outgroup = 0/Failure in Binomial terms) onto desirability in a binomial GLMM, while desirability was estimated with varying slopes across subjects, while varying intercepts were estimated across subjects and traits.

Association between Self-Evaluations and Group Predictions. We additionally sought to replicate traditional tests of self-anchoring by relating initial self-evaluations to group classifications as a one-to-one comparison. Self-evaluations were estimated with freely varying slopes across subjects, and varying intercepts across subjects and traits. Notably, this approach does result in a considerable loss of usable data for the model, given that only 90 traits are self-evaluated on but all 148 traits are observed during the generalization phase (approximately 58 traits observed during generalization with missing data for regressor).

Similarity-Based Generalization to Group. In order to test the extent to which people generalize from the self-concept to the ingroup across semantically similar traits, we tested whether a trait's Similarity-to-self predicts ingroup classifications more than outgroup classifications. We focused on whether a particular relational measure,

[1] A two-way, average, consistency ICC revealed good reliability for the normatively rated social desirability of positive traits, ICC(3, k) = .682 [.601, .753].

Similarity-to-self, predicts ingroup classifications. *Similarity-to-self* was estimated with varying slopes for subjects, and varying intercepts for subjects and traits.

We further controlled for desirability in separate models, to account for the general tendency to classify desirable traits as characteristic of the ingroup. We additionally were interested in controlling for people's average beliefs for the ingroup for each. For instance, some traits may be perceived as more characteristic of overestimators or underestimators on average. As such, we average the choices for each trait across all participants to estimate how typical a trait is considered of the ingroup on average. In a separate model, we control for this effect to determine the extent to which the classifications are due to a trait merely being perceived as characteristic of the ingroup on average. Finally, we also compute an average for each trait of the self-descriptiveness across all participants, and control for this in a separate model to evaluate whether more generally descriptive traits explain the effect, rather than more Similar-to-self traits specifically.

Additionally, to test whether people generalize beyond only the "trained" evaluated traits, we tested whether the effect of Similarity-to-self on group predictions generalizes beyond merely the traits repeated from the training phase. To model this, we tested an interaction between novelty– a dummy coded categorical factor denoting whether the traits were observed during learn or not)-- with Similarity-to-self, to test whether the slopes for similarity-to-self differ depending on whether the trait was novel

or repeated. We estimated novelty with varying intercepts and slopes for subjects, and varying slopes for traits. Given that we were primarily interested in whether the effect of similarity-to-self was as strong for novel traits as repeated traits, we conducted an equivalence test (Lakens et al., 2018; Lüdecke et al., 2021) to test whether the 95% CI of the interaction was largely overlapping with the region of practical equivalence (i.e., the region around 0).

Additionally, in prior work we found that individuals resist updating self-beliefs from social feedback (J. Elder, Davis, et al., 2023b; J. Elder et al., 2022c) and across time (J. Elder et al., 2022a) for traits with more downstream implications. We sought to further explore whether there may be trait-by-trait differences in tendencies to self-anchor as a function of degree centrality. We model an interaction of indegree and outdegree centrality with Similarity-to-self to determine if the tendency to self-anchor and generalize from the self-concept to the group differs depending on the number of trait dependencies.

We implement other relational measures, but for brevity, we focus only on the Similarity-to-self measure. These measures largely provide similar inferences, but other possible measures using the network include using the average self-evaluation across a trait's immediate neighbors, the dot-product of response categories with response probabilities to estimate expected rating ($1*P_1 + 2*P_2 + 3*P_3 + 4*P_4 + 5*P_5 + 6*P_6 + 7*P_7$), or cross-validated predicted self-evaluations trained on the self-evaluation data.

Association Between the Metacontrast Ratio and Self-Evaluations. We tested the extent to which the Similarity-to-ingroup over Similarity-to-outgroup (i.e., *Metacontrast Ratio*) is associated with the self-descriptiveness of traits. We log transform the metacontrast ratio (ratios are better behaved in regression analyses when log transformed) and use it as a predictor of people's initial self-evaluations. We regress initial self-evaluations onto the metacontrast ratio in a CLMM. Notably, while self-evaluations occur before group classifications, the self-evaluations in this model are estimated as the outcome. We estimate the metacontrast ratio with varying slopes for subjects, and estimate varying intercepts for subjects and traits. In a second model, we further split up Similarity-to-ingroup and Similarity-to-outgroup as separate predictors to consider the relative contributions of each, with varying slopes for subjects, and varying intercepts for subjects and traits.

Correlates with Trait Segregation. We were interested in measuring how individual differences in the tendency to segregate the ingroup and outgroup classifications from one another might correlate with other individual differences. To evaluate this, we use the network homophily measure of Trait Segregation, and correlate it with all of our individual differences self-report measures. For our correlations, due to the number of comparisons, we omit p-values and focus strictly on effect sizes and CIs.

Correlates with Projection Rate. One advantage of computational modeling is the ability to relate person-level computational parameters to other individual differences

measures in order to draw inferences (Daw, 2011), in this context about how people generalize from the self to the group. With this in mind, we use the *projection rate* and correlate it with all of our individual differences self-report measures. We did not formally preregister these analyses given that the computational modeling approach emerged after the collection and analysis of data, but it would be consistent with theory to find that the projection rate is more associated with social identification, self-prototypicality, and need for cognition (van Veelen, Otten, et al., 2016). As exploratory analyses, we also measure the correlations between all other computational parameters and self-report measures.

Open Science

Analytic code and materials for all studies, and pre-registration for this study, can be found at the following link:

https://osf.io/tc94q/?view_only=7c66bd5dd475420e91b56536e2487352

Results

People Classify Desirable Traits as Ingroup Characteristic

As a first sanity check test, we tested whether individuals are likely to classify more desirable traits as characteristic of the ingroup. Indeed, we found support for this (β = .074, *OR* = 1.08, $CI_{95\%}$ = [-.01, .157], *pd* = 95.5%, *p* = .09, *ps* = 34.9%, *pl* = 0%), providing weak evidence that individuals are motivated to view their social group positively, even a minimal group. This provides evidence of ingroup enhancement for minimal groups and replicates prior work (Figure 7A).

People Classify Self-Descriptive Traits as Ingroup Characteristic

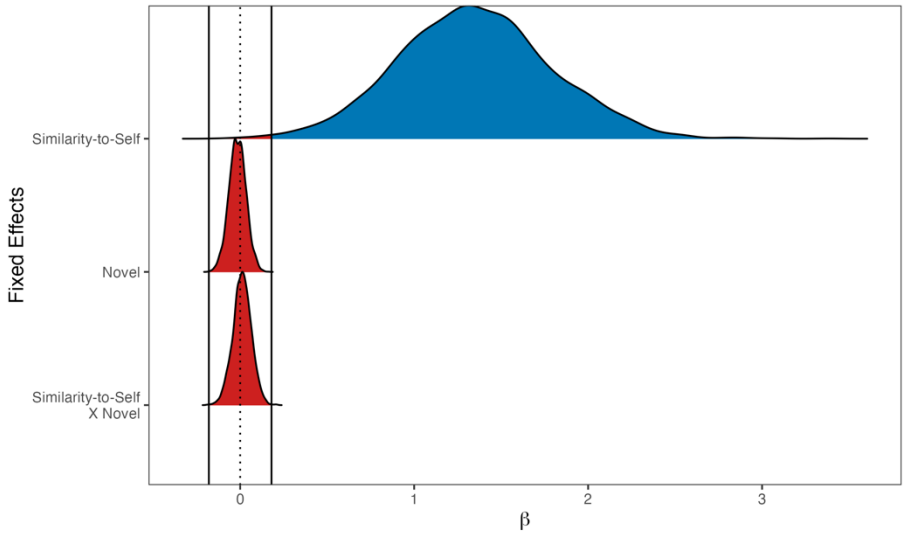
We next sought to implement the more traditional test of self-anchoring, which determines whether self-evaluations are associated with group-evaluations. We found support for this (β = .339, *OR* = 1.404, $CI_{95\%}$ = [.211, .466], *pd* = 100%, *p* = .0, *ps* = 100%, *pl* = 0.05%), and found that more self-descriptive traits are more likely to be classified as characteristic of the ingroup. This effect was robust even while controlling for the desirability of the traits. This provides evidence of self-anchoring via repeated evaluations (Figure 7B).

People Classify Similar-to-Self Traits as Ingroup Characteristic

We next sought to move beyond traditional tests of self-anchoring, and to provide a more mechanistic test of generalization, and to characterize whether people generalize to the ingroup across similar traits, rather than only the same traits. We find that traits that are more similar to oneself are more likely to be classified as the ingroup (β = 1.394, *OR* = 4.033, $CI_{95\%}$ = [.591, 2.260], *pd* = 99.93%, *p* = .0015, *ps* = 99.85%, *pl* = 98.35%). This reflects that people are not only likely to classify a self-descriptive trait (e.g., outgoing) as characteristic of one's ingroup, but also similar traits to the self-descriptive trait (e.g., sociable, fun, and witty) as characteristic of one's ingroup. Moreover, while the traditional test of self-anchoring using one-to-one comparisons revealed a small probability of being a large effect, the similarity-based test revealed a much larger probability of being a large effect, suggesting that the similarity-based approach better reflects how people engage in these ingroup trait inferences (Figure 7C).

We further test whether more Similar-to-self traits are likely to be classified as characteristic of the ingroup, even if they were never self-evaluated on during the training phase in the first place. An equivalence test provided support for the null (ROPE = [-.18, .18], $CI_{95\%}$ = [-.09, .11], Inside ROPE = 100%), suggesting that there were no differences in the Similarity-to-self slopes for novel and repeated traits (Figure 6).

Figure 6. Depiction of probability of significance and equivalence test in Study 1.

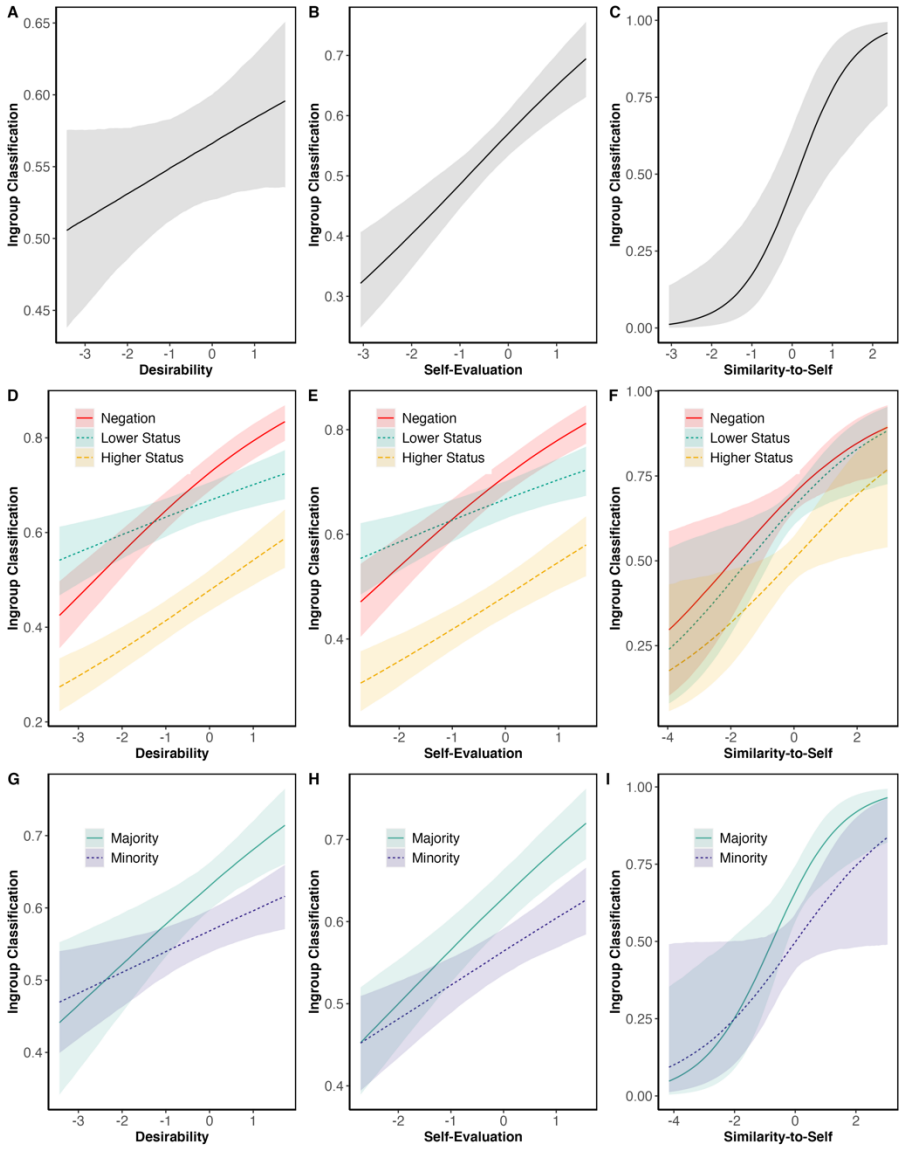


Note. The solid vertical line denotes the threshold for "significance" (i.e., $\beta > |.18|$) of the estimate. Red denotes the proportion of the distribution that is roughly equivalent to 0 due to being within the Region of Practical Equivalence (ROPE), while blue denotes the proportion of the distribution that is significant.

In prior work, we found that people were less likely to change from social feedback (J. Elder, Davis, et al., 2023b; J. Elder et al., 2022c) and across time (J. Elder et al., 2022a) for traits with more implications. Traits with more semantic implications thus appear to be important for learning and self-concept change, but what role do they play in self-projection to the ingroup? We did not find evidence that the effect of similarity-to-self depends on outdegree centrality (β = .034, *OR* = 1.035, $CI_{95\%}$ = [.591, 2.260], *pd* = 83.43%, *p* = .332, *ps* = 6.32%, *pl* = 0%). We also conducted this test for indegree centrality and found weak evidence that generalization using a trait's similarity-to-self was stronger for higher indegree centrality traits (β = .046, *OR* = 1.048, $CI_{95\%}$ = [-.006, .102], *pd* = 96.00%, *p* = .080, *ps* = 5.98%, *pl* = 0%). Findings provide insufficient

evidence to suggest that generalization depends on a trait's indegree or outdegree centrality.

Figure 7. Predictors of ingroup classification: Desirability, self-evaluations, and similarity-to-self.



Note. Plotted using marginal effects with 1.96 +/- SEs. Top row is Study 1. Middle row is Study 2. Bottom row is Study 3. Desirability as predictor is left column, self-evaluations is middle column, similarity-to-self is right column. The x-axis is the Z-scored predictor, the y-axis is the probability of the outcome, ingroup classification. The colors denote different conditions for each study.

Self-Uncertainty Predicts Less Ingroup Classification

The previous results suggest that participants consider how similar a trait is to prior self-evaluations in discerning whether to generalize these self-evaluations to the ingroup. We next use an information-theoretic measure of uncertainty to predict ingroup classifications, and find that traits which were associated with greater uncertainty were less likely to be classified as characteristic of the ingroup (β = -.739, *OR* = .478, $CI_{95\%}$ = [-1.389, -.010], *pd* = 98.73%, *p* = .0255, *ps* = 97.53%, *pl* = 72.80%). This provides further support for the similarity-based mechanisms underlying self-concept generalization to the ingroup, as to the extent that traits are similar to traits that received a variety of different self-evaluations, there is greater uncertainty and thus less ability to classify the trait as belonging to the ingroup.

Projection to the Ingroup, Rather than Rejection of the Outgroup

To further distinguish the role of either projection to the ingroup, rejection of the outgroup, or some combination of both processes, we next investigated the extent to which self-evaluations are associated with similarity-to-ingroup and similarity-to-outgroup (as defined by traits classified as characteristic of ingroup or outgroup respectively). Using a CLMM, we regressed ordinal Likert self-evaluations onto each trait's log Metacontrast Ratio, denoted by a self-evaluated trait's summed similarity to all ingroup classified traits over its summed similarity to all outgroup classified traits. This thus tests the extent to which trait self-evaluations are associated with the aggregate similarity of a trait to all trait's that the participant will classify as characteristic of the ingroup or characteristic of the outgroup. A trait's Metacontrast Ratio was strongly

associated with a trait's self-descriptiveness (β = .682, *OR* = 1.978, $CI_{95\%}$ = [.394, .892], *pd* = 100%, *p* = .000, *ps* = 100%, *pl* = 99.38%), such that as a trait's Metacontrast Ratio is higher (similarity to ingroup or dissimilarity from outgroup), people previously evaluated more self-descriptively on the trait. To further interrogate the relative contributions to this process, we separated the similarity into separate regressors and found that summed ingroup similarity was strongly associated with self-evaluations (β = .591, *OR* = 1.806, $CI_{95\%}$ = [.384, .812], *pd* = 100%, *p* = .000, *ps* = 100%, *pl* = 99.60%), whereas summed outgroup similarity was weakly negatively associated with self-evaluations (β = -.162, *OR* = .850, $CI_{95\%}$ = [-.341, .012], *pd* = 95.55%, *p* = .069, *ps* = 90.13%, *pl* = 6.65%). This may reflect that it is primarily the drive to achieve similarity with the ingroup, rather than to reject or be repulsed by the outgroup, that accounts for the association between self-evaluations and group classifications. However, there is some weak evidence that a trait's similarity to the outgroup is associated with less self-descriptiveness, providing potential evidence of outgroup repulsion.

Correlates of Trait Segregation

We next explored the individual differences that are associated with a greater tendency to segregate the traits belonging to the ingroup and outgroup. We find that individuals higher in self-esteem (*r* = -.25, CI = [-.44, .01) and independence (*r* = -.20, *CI* = [-.41, .05]) segregate traits less for their ingroup and outgroup, whereas those higher in interdependence (*r* = .17, CI = [-.06, .40]) and need to belong (*r* = .26, CI = [.04, .47]) segregate traits more for their ingroup and outgroup. It may be that individuals higher in the need to belong and interdependence are more motivated to "carve" differences

between the ingroup and outgroup, whereas those higher in self-esteem experience less of a need to divide groups into sharper categories.

Generalization Model

Model Performance. One advantage of computational modeling is to provide insight into the model that best depicts a given cognitive process based on its success relative to alternative models. Here, we find that *self-projection model*, whereby self-evaluations are converted to group beliefs prior to being projected via similarity, outperforms the bias only model.

The self-projection model converged such that there was only one Rhat values above 1.01, and none below an Rhat of 1.05, with a maximum of Rhat = 1.011.

Table 1. Model comparisons for computational models across studies.

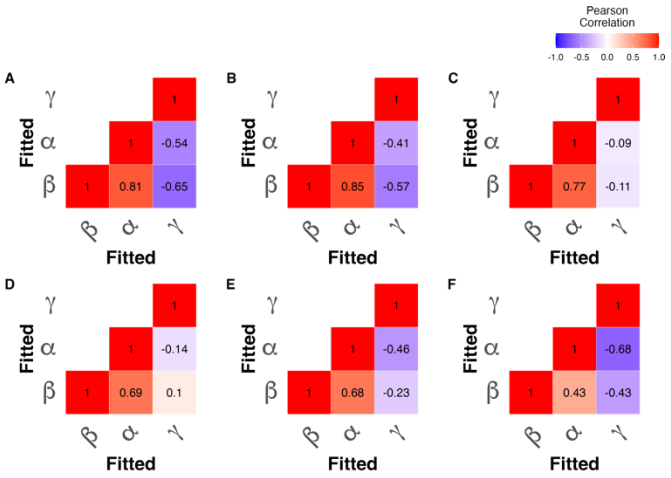
Model	LOO	LOO-SE	LOO-Diff (SE-Diff)	No. Est. Parameters
Study 1: Minimal				
Self-Projection	-5604.7	103.6	--	3
Bias	-5852.8	83.9	-248.1	1
Study 2: Higher Status				
Self-Projection	-9047.9	112.8	--	3
Bias	-9462.0	91.4	-414.1	1
Study 2: Lower Status				
Self-Projection	-8206.3	145.3	--	3
Bias	-8232.8	141.1	-26.5	1
Study 2: Negation				
Self-Projection	-7769.3	241.8	--	3
Bias	-7849.0	238.4	-79.7	1
Study 3: Majority				
Self-Projection	-11503.8	207.9	--	3
Bias	-11714.2	198.5	-210.4	1
Study 3: Minority				
Self-Projection	-13060.4	157.6	--	3
Bias	-13384.2	138.8	-323.8	1

Note. LOO = sum PSIS-LOO, approximate leave-one-out cross-validation (LOO) using Pareto-smoothed importance sampling (PSIS); LOO-SE = Standard error of PSIS-LOO; LOO-Diff (SE-Diff) = Difference in expected predictive accuracy (PSIS-LOO) for all models from the model with the highest PSIS-LOO; No. Est. Parameters = number of estimated parameters in the model.

Parameter Recovery. In order to determine whether parameters from the computational model were identifiable and could be recovered, we simulated data across N = 250. In order to resemble how participants self-evaluate similarly on similar traits in the training data, we randomly sampled participants' real training data and only simulated the generalization phase classifications as a function of similarity to the training trait self-evaluations. We found that the projection rate was recoverable (r = .76), the bias parameter was modestly recoverable (r = .40), and the temperature parameter was not recoverable at all (r = .01). The low recoverability of the temperature parameter is potentially concerning but we do not utilize the temperature parameter for individual-level inference (see Figure 8 for correlations among fitted parameters and Figure 9 for parameter recovery correlations).

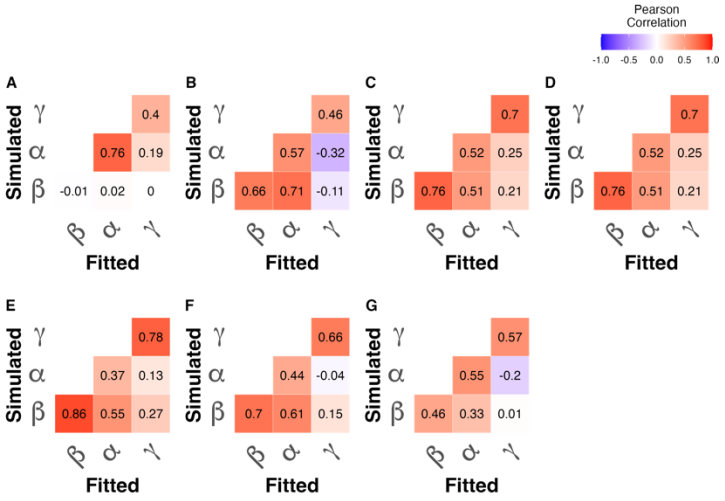
We next performed parameter while retaining the original covariance structure among parameters. To do so, we simulated behavior using the original participants' fitted parameters. We then fit parameters to this simulated behavior to determine the extent to which the original participants' parameters could be recovered. We found that the temperature parameter (r = .66), projection rate (r = .57), and bias parameter (r = .46) were modestly recoverable.

Figure 8. Correlations among fitted parameters.



Note. (A) Study 1 minimal group, (B) Study 2 higher status outgroup condition, (C) Study 2 lower status outgroup condition, (D) Study 2 negation outgroup, (E) Study 3 majority outgroup condition, (F) Study 3 majority outgroup condition.

Figure 9. Parameter recovery



Note. (A) Randomly generated parameters, (B) Study 1 parameters (C) Study 2 higher status condition parameters, (D) Study 2 lower status condition parameters (E) Study 2 negation condition parameters, (F) Study 3 majority condition parameters, (G) Study 3 minority condition parameters.

Correlations in Parameters. Another advantage of computational modeling is the ability to draw inferences based on the correlations between computational parameters and individual differences. We associated the computational parameters with an array of other individual differences. We found weak evidence of a positive association between projection rate and social identification (r = .10, $CI_{95\%}$ = [-.15, .32]), self-concept clarity (r = .12, $CI_{95\%}$ = [-.12, .35]), and self-prototypicality of the group (r = .08, $CI_{95\%}$ = [-.15, .32]), and we found weak evidence of a negative association between projection rate and self-esteem (r = .15, $CI_{95\%}$ = [-.39, .08]). While the effect sizes are small, the findings provide some interesting suggestive evidence that the extent to which people convert their self-beliefs into ingroup beliefs is a reflection of how strongly identified they are with their social groups (M. Cadinu & Rothbart, 1996; van Veelen, Otten, et al., 2016), and the extent to which they rely on group identification to fulfill positivity needs (R. Brown, 2000; Hogg & Abrams, 1988; Tajfel, 1978). It is also important to note that these weak associations are identified for minimal groups with which the participants have no ostensible prior attachments, and thus effect sizes should be expected to be attenuated for these individual differences associations.

For the bias parameter, we found that more independent individuals were more biased towards outgroup classifications regardless of self-beliefs (r = -.25, [-0.46, -0.01]), while more interdependent individuals were more biased to ingroup classifications (r = .19, [-0.06, 0.40]). Additionally, individuals with higher self-esteem (r = -.12, $CI_{95\%}$ = [-0.34, 0.13]), self-concept clarity (r = -.15, $CI_{95\%}$ = [-0.40, 0.08]), and need for cognition (r = -.12, $CI_{95\%}$ = [-0.36, 0.11]) are biased towards outgroup classifications while

individuals higher in dialectical self-views (r = .16, $CI_{95\%}$ = [-0.07, 0.39]) and need to belong (r = .13, $CI_{95\%}$ = [-0.11, 0.36]) are biased towards ingroup classifications. Interestingly, findings suggest that individuals who are more motivated towards social groups may be more likely to classify traits as characteristic of the ingroup, regardless of self-beliefs.

Discussion

We find support for similarity-based mechanisms for self-concept generalization to the ingroup over the outgroup, under minimal conditions. Importantly, the effect of the similarity-based self-concept generalization is much stronger than mere self-evaluation as is typically used to characterize self-anchoring or self-projection. Moreover, this generalization is robust across both novel and repeated traits, discarding the potential interpretation that self-anchoring is a function of repetition effects (Unkelbach et al., 2019; Unkelbach & Rom, 2017). This provides initial evidence that people use relational similarity to infer from one's self-concept to other attributes what may be characteristic of one's ingroup. Our metacontrast effects provide evidence that self-evaluations are positively associated only with similarity-to-ingroup, while similarity-to-outgroup is not associated with self-evaluations, suggesting that people may be less repelled from outgroup similarity than they are attracted to achieving ingroup similarity. Further, our design provides evidence that self-anchoring is accentuated by contrasting against one's outgroup. Finally, our computational model provides insight into how people may convert self-beliefs into ingroup-beliefs, suggesting that individuals higher in social identification and self-prototypicality of the ingroup are more extreme in their tendencies

to convert self-beliefs to ingroup-beliefs. However, it is unclear to what extent features of the outgroup relative to the ingroup may motivate stronger self-anchoring and tendencies to generalize from the self.

Study 2: Self-Anchoring Based on Relative Status of University Groups

In the next study, we sought to examine what diagnostic features of an ingroup-outgroup contrast may promote self-anchoring to the ingroup. Specifically, we attempt to address this using real groups that differ in perceived social status, focusing on different universities. The student body at the research institute where this data was collected served as the ingroup, while a higher status university and lower status university in the same Southern California region of the U.S. were treated as two of the outgroup contrasts. An additional outgroup contrast was tested, which was simply the negation of the ingroup– "Not Ingroup University"-- drawing upon concept learning designs which use either multiple concepts or the negation of a concept as an alternative option (Zeithamova et al., 2008). Ingroup versus Not Ingroup may lead to greater attention to the characteristics of the Ingroup specifically, whereas Ingroup versus Outgroup should lead to more contrastive effects and greater attention to the differences (Davis & Love, 2010). More simply, ingroup versus outgroup implies that the participant should attend to what separates an ingroup from an outgroup, while ingroup versus not ingroup on the other hand implies that the participant should attend to what is unique about the ingroup itself. People focus accordingly and treat ingroup versus not ingroup as an opportunity to classify primarily based on ingroup, and less about what is not ingroup.

This design allows us to investigate the contrast for which self-anchoring will be strongest, presumably due to perceived differences in status. A further advantage to this design is that there has been relatively scarce evidence of self-anchoring being applied to real groups, beyond minimal groups (Otten & Epstude, 2006; Riketta & Sacramento, 2008; van Veelen et al., 2011, 2013a), and this allows us to establish further evidence of self-anchoring in real groups that are also relatively unclearly defined with fewer diagnostic or normatively defined characteristics than other social identities. We propose competing hypotheses: (a) Individuals may be more motivated to self-anchor when contrasted against a lower status outgroup which they see themselves as highly differentiated from due to a positive self-concept or a desire to achieve a positive self-concept (Tajfel, 1978), (b) people may self-anchor more when compared against a higher status outgroup due to a motivation to justify and maintain status differences (Jost & Banaji, 1994), or (c) people may self-anchor regardless of the outgroup comparison, which may be supported by the fact that positive ingroup evaluations rely more so on positive associations with the self rather than explicit comparisons with the outgroup (Brewer, 1999).

Methods

Participants

We recruited 337 participants and excluded 54 participants to arrive at a final sample of N = 283. We excluded any participants who self-reported not taking the task seriously (less than '4' to "To what extent did you take this task seriously?"; N = 2), who self-reported that their data is unusable ("No" to "Did you understand the task and

respond truthfully and meaningfully enough that your data is usable?"; N = 4). We additionally excluded participants if they exhibited behavior during the task that reflected careless responding, with exclusionary criteria including if over 80% of their self-evaluations were identical, if over 95% of their group classification were identical, or if over 40% of their behavioral responses were missing from either part of the task (N = 47).

Participants (N = 283) were native English-speaking university students (65.56% Cisgender Female, 33.33% Cisgender Male, 1.11% Nonbinary; M_{Age} = 19.52, SD_{Age} = 1.82, $Range_{Age}$ = [18, 36]), and were 4.43% White/Caucasian, 4.80% Black/African-American, 33.21% Mixed/Other, 9.96% Asian, 32.10% Hispanic/Latino, 0.37% Native-American, 9.59% Indian/South Asian, and 5.54% Middle Eastern/North African.

We conducted a power analysis in *simr*, tripling the size of the dataset and labeling each different dataset as a different level of condition. We used a similarity-based predictor and assigned one level β = -.35, and the other an interaction β = .35, in comparison to a dummy-coded third level. Power curve analysis from 50 to 225 revealed that above 80% power was achieved by 200 participants for testing the effect of one of the slopes relative to the reference group. We over-recruited due to the large number of exclusions, and also due to the fact that reliability of the measures can attenuate sensitivity (Blake & Gangestad, 2020; Spearman, 1904), which means conventional power analyses that disregard measurement error are "best-case scenarios".

Design

The design was identical to the previous study's design. However, no minimal group assignment was involved. Rather, roughly 33% of participants were randomly assigned to the Negation outgroup comparison, roughly 33% of participants were randomly assigned to the Higher Status outgroup comparison, and roughly 33% of participants were randomly assigned to the Lower Status outgroup comparison. The Higher Status and Lower Status outgroups in this study were two local universities to the ingroup university which participants attended, but which varied in relative prestige and status.

Measures

The same measures were used as in Study 1, except for some additional self-report individual differences measures. We collected the perceived warmth for each group using a Feeling Thermometer ranging from 0 to 100 (Iyengar et al., 2019) and the perceived social status of each group using the MacArthur Social Status Ladder (Adler et al., 1994). We also collected measures of perceptions of Self-Group Overlap between participants, the ingroup university, and the two outgroup universities involved in the study (Schubert & Otten, 2002). We altered our measure of social identification from the single-item self-report to the Multidimensional Group Identification Scale (Leach et al., 2008) for greater granularity in the measurement of social identification processes. The MGIS contains subscales for Solidarity, Satisfaction, and Centrality, which are further averaged to reflect Group-Level Self-Investment. It also contains subscales for Individual Self-Stereotyping and Ingroup Homogeneity, which are averaged to reflect Group-Level

Self-Definition. Of particular interest in the "Individual Self-Stereotyping" subscale, which contains the questions, "I have a lot in common with the average [In-group] person," (Spears et al., 1997) and "I am similar to the average [In-group] person," (Doosje et al., 1995; Spears et al., 1997) which reflect perceptions of self-similarity with the ingroup. We average all the measures in the MGIS to reflect overall social identification. We also measure Collective Self-Esteem (Luhtanen & Crocker, 1992).

Planned Analyses

Differences Across Universities in Perceived Warmth and Status. As a manipulation check to test whether the different universities are indeed perceived as qualitatively different, we modeled the group differences in terms of perceived social status (as defined by MacArthur Social Status Ladder) and the participants' feelings of warmth towards them (as defined by the Feelings Thermometer). To do so, we implemented a mixed-effects model with subject as a random factor, and university as a dummy coded categorical factor with varying intercepts across subjects. For the model with the Feelings Thermometer as the outcome, we used the *ordbetareg* package (Kubinec, 2022) to implement a ordered beta regression model for the 0 to 100 slider scale outcome with upper and lower boundaries (scaled by 100 for beta distributed data). For the model with Social Status as the outcome, we implemented a Cumulative Link Logistic model.

Self-Anchoring Moderated by Outgroup Comparison. We implemented identical analyses as in Study 1. However, we modeled an interaction between condition and our trait-level predictors of ingroup classification, to test whether their effects differ

between the conditions. Condition was treated as a dummy coded factor with the Negation level as the reference group, as it is assumed to be akin to a control group with no explicitly diagnostic or differentiating features about the status of the comparison option. We allowed the effect of condition to vary across traits.

As a further exploratory test, we modeled a three-way interaction between Similarity-to-self, the novelty factor, and the condition factor, to test whether generalization to unobserved traits differs across outgroup comparisons.

We additionally tested whether the outgroup comparison moderated the effect of summed similarity of trait self-evaluations to ingroup and outgroup choices. First, we estimated a model where self-evaluations were regressed the metacontrast ratio, condition, and the interaction between the two, while the effect of the metacontrast ratio was allowed to vary across subjects and the effect of condition was allowed to vary across traits. Nest, the effect of Similarity-to-ingroup and Similarity-to-outgroup on were modeled, while also estimating both of their interactions with condition. Similarity-to-ingroup and Similarity-to-outgroup were estimated with varying slopes for subjects, condition was estimated with varying slopes for traits, and both subjects and traits were estimated with varying intercepts.

Correlates and Differences in Trait Segregation. We additionally performed an exploratory test whether there are differences between contrasts (each level of the different conditions) in subject-level Trait Segregation. To test this, we implemented a Bayesian regression model predicting Trait Segregation (i.e., homophily in group

classifications for traits) by the dummy coded categorical condition factor, with the "Higher Status" level treated as the reference group.

Results

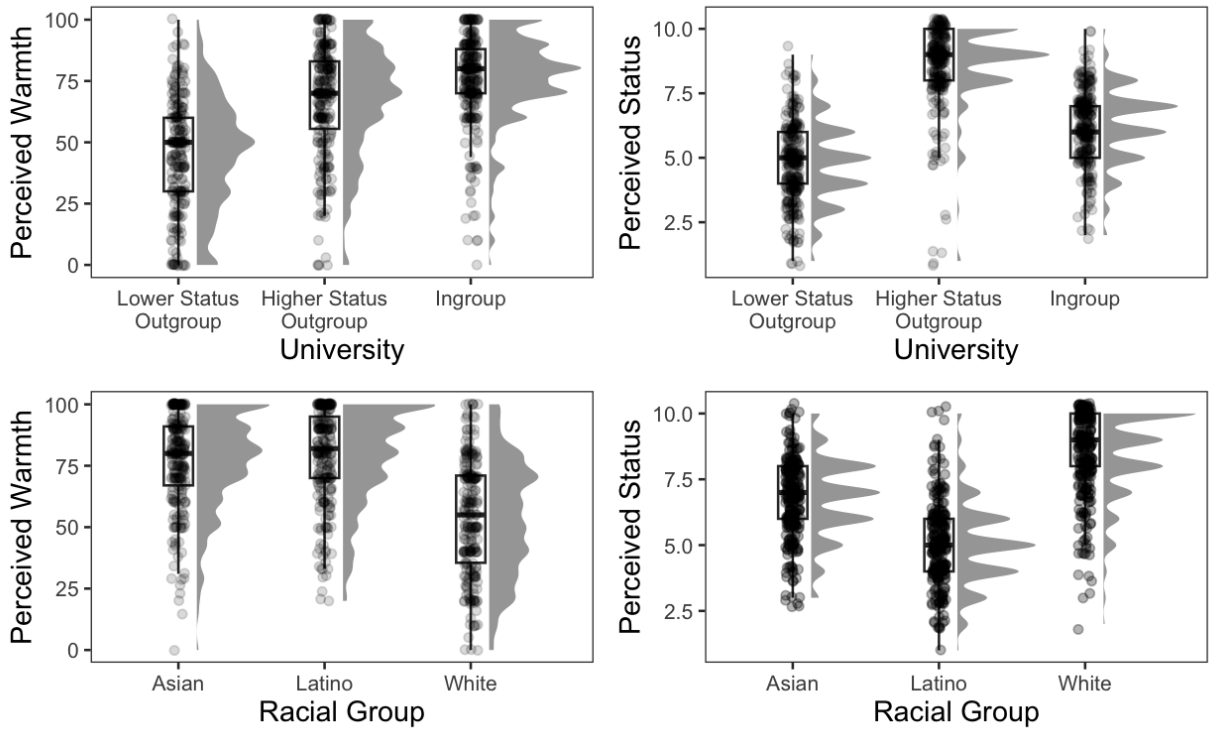
Differences Across Universities in Perceived Warmth and Status

First, to explore whether the assumptions of differences in perceived social status between universities of our design were well-founded, we tested for differences in perceived social status across the universities used in the study. Indeed, we found that compared to the ingroup university (M = 6.22, SD = 1.43) the ostensibly higher status university was evaluated as higher status (β = 3.718, *OR* = 41.212, $CI_{95\%}$ = [3.275, 4.178], *pd* = 100%, *p* = .0, *ps* = 100%, *pl* = 100%; M = 8.58, SD = 1.55) and the ostensibly lower status university was evaluated as lower status (β = -1.835, *OR* = .160, $CI_{95\%}$ = [-2.184, -1.522], *pd* = 100%, *p* = .0, *ps* = 100%, *pl* = 100%; M = 4.74, SD = 1.51). This provides a manipulation check and support for the premise that these outgroups represent relative differences in perceived social status.

Next, to explore whether universities are differently evaluated in terms of feelings of positivity towards them, we compared differences in perceived warmth across universities within each participant. We find that the participants perceived their own university most warmly (M = .76, SD = .18) and that the higher status university was perceived less warmly (β = -.297, $CI_{95\%}$ = [-.440, -.155], *pd* = 100%, *p* = .0, *ps* = 99.95%, *pl* = 47.98%; M = .68, SD = .22), and the lower status university was perceived much less warmly (β = -1.121, $CI_{95\%}$ = [-1.269, -.973], *pd* = 100%, *p* = .0, *ps* = 99.95%, *pl* = 100%; M = .45, SD = .23). Thus, participants appear to perceive their own university most

positively, the higher status university less positively (but still quite positively), and the lower status outgroup university least positively (see Figure 10 for perceived status and positivity among university groups).

Figure 10. Raincloud plot depicting differences in status and positivity.



Note. University groups (top row) and racial groups (bottom row).

People Classify Desirable Traits as Ingroup Characteristic

We first sought to replicate the prior finding of whether people would project desirability onto the ingroup. Indeed, this was supported again (β = .261, *OR* = 1.299, $CI_{95\%}$ = [.190, .330], *pd* = 100%, *p* = .0, *ps* = 100%, *pl* = 0%), such that people classified more desirable traits as more ingroup typical. We next tested whether this effect differed depending on the outgroup that they compared their ingroup against. Interestingly, this

tendency to project desirability onto the ingroup was greatest for the Negation comparison. There was insufficient evidence of differences for the higher status comparison (β = -.081, *OR* = .892, $CI_{95\%}$ = [-.232, .070], *pd* = 84.03%, *p* = .3195, *ps* = 45.03%, *pl* = 0%). However, for the lower status comparison (β = -.228, *OR* = .922, $CI_{95\%}$ = [-.364, -.090], *pd* = 99.99%, *p* = .0015, *ps* = 97.45%, *pl* = 0%) the tendency to classify desirable traits as characteristic of the group was weaker relative to the Negation comparison (Figure 7D). Interestingly, people may engage in less ingroup enhancement when they are comparing against a lower status university, which may be because there is less need to enhance the ingroup or derogate the outgroup if the outgroup is perceived as lower status already.

People Classify Self-Descriptive Traits as Ingroup Characteristic

We again implemented the traditional test of self-anchoring which strictly examines one-to-one comparisons of self- and ingroup-evaluations. Indeed, we found that the more self-descriptive a given trait is, the more likely it is to be classified as ingroup (β = .270, *OR* = 1.310, $CI_{95\%}$ = [.217, .330], *pd* = 100%, *p* = .0, *ps* = 100%, *pl* = 0%), replicating prior findings. Interestingly, we found that in comparison to participants contrasting their ingroup against the Negation, participants contrasting against a lower status university (β = -.136, *OR* = .873, $CI_{95\%}$ = [-.276, .004], *pd* = 97.18%, *p* = .0565, *ps* = 73.18%, *pl* = 0%) or higher status university (β = -.167, *OR* = .846, $CI_{95\%}$ = [-.305, -.030], *pd* = 99.20%, *p* = .0160, *ps* = 86.95%, *pl* = 0%) exhibited less of a tendency to self-project (Figure 7E). It may be that the tendency to self-project to the ingroup on repeated

evaluations is greater when attending to the unique characteristics of the ingroup rather than the differences between the ingroup and an outgroup.

People Classify Similar-to-Self Traits as Ingroup Characteristic

We again tested whether people will classify traits that are similar to self-descriptive traits as characteristic of the ingroup rather than the outgroup. Indeed, we found that traits that are higher in similarity-to-self are more likely to be classified as ingroup (β = .517, *OR* = 1.677, $CI_{95\%}$ = [.287, .750], *pd* = 99.95%, *p* = .001, *ps* = 99.93%, *pl* = 40.28%). Again, the effect of this similarity-to-self is descriptively larger than the conventional manner of measuring self-anchoring by comparing repeated trait evaluations, reflecting added value to considering the similarity-based mechanisms underlying this generalization process. An equivalence test again suggested that the effect of similarity-to-self for novel traits and previously self-evaluated traits are equivalent to one another (ROPE = [-.18, .18], Inside ROPE = 100%, $CI_{95\%}$ = [-.06, .03]). This suggests that the tendency to self-project one's own attributes to the group, even across different but related traits, also occurs for real groups such as universities.

We further explored whether this tendency to use similarity-to-self in order to generalize to the ingroup differs depending on the relative status of the outgroup that the ingroup is compared against. Compared to the negation condition, there was insufficient evidence to suggest that the effect of similarity-to-self on ingroup classifications differed for participants comparing their university to the higher status university (β = -.042, *OR* = .959, $CI_{95\%}$ = [-.479, .409], *pd* = 56.98%, *p* = .861, *ps* = 41.13%, *pl* = 0.95%) or for participants comparing their university to the lower status university (β = .024, *OR* =

1.024, $CI_{95\%}$ = [-.427, .470], *pd* = 54.28%, *p* = .9145, *ps* = 38.60%, *pl* = 1.24%). While there was evidence of outgroup comparison contributing to differences in the effect of self-evaluations and desirability on ingroup classifications, it appears that the effect of Similarity-to-self does not depend on the outgroup comparison (Figure 7F).

Lastly, we again tested whether the centrality of a trait contributes to how readily it is generalized to the ingroup. Participants exhibited a slightly stronger tendency to generalize to the ingroup using similarity-to-self for higher relative to lower outdegree traits (β = .075, *OR* = 1.078, $CI_{95\%}$ = [.039, .111], *pd* = 100%, *p* = .0, *ps* = 19.23%, *pl* = 0%). We found weak evidence for generalization depending on indegree centrality (β = .020, *OR* = 1.020, $CI_{95\%}$ = [-.003, .045], *pd* = 95.98%, *p* = .0805, *ps* = 0%, *pl* = 0%). Interestingly, we did not find this evidence for the role of outdegree centrality in supporting generalization in Study 1, but that may be due to a smaller sample size in Study 1. We provide a weak replication for the finding that Similarity-to-self is more predictive of ingroup classifications for higher indegree traits. Higher indegree centrality traits may provide greater inferential evidence with which to engage in generalization about one's group, as people may engage in more inferences about traits for which there are more causes.

Self-Uncertainty Predicts Less Ingroup Classification

We again explore whether self-evaluative uncertainty predicts the likelihood of ingroup classifications. We find that traits which were associated with greater uncertainty were weakly negatively associated with ingroup classifications (β = -.182, *OR* = .834, $CI_{95\%}$ = [-.394, .046], *pd* = 93.88%, *p* = .1225, *ps* = 78.00%, *pl* = 0.08%). While a weaker

effect than in Study 1, this again provides consistent evidence that to the extent that there a trait's similarity to prior trait self-evaluations provides less certainty in inferences, people are less likely to classify the trait as characteristic of the ingroup.

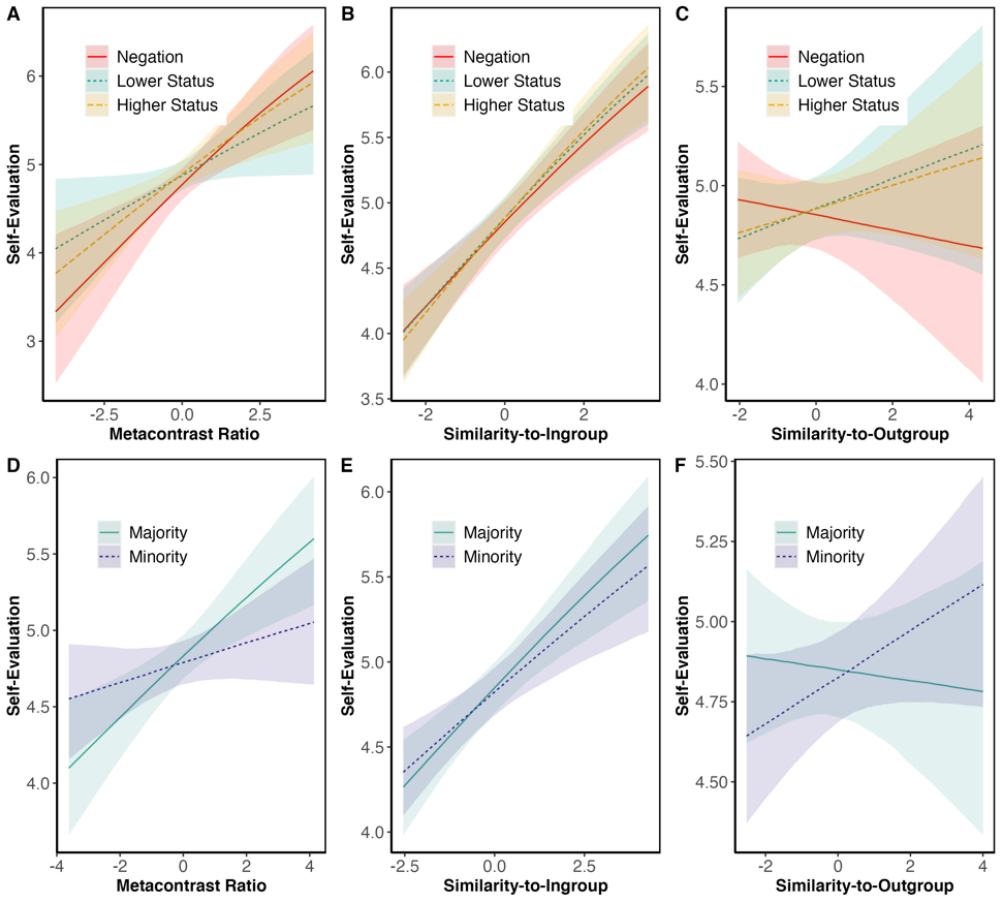
Projection to the Ingroup, Rather than Rejection of the Outgroup

To further determine the contribution of ingroup similarity or outgroup repulsion, we examined the association between initial self-evaluations during learning with the summed similarity to ingroup choices and the summed similarity to outgroup choices. As was the case for minimal groups, we found that the metacontrast ratio was again strongly associated with self-evaluations (β = .399, OR = 1.490, $CI_{95\%}$ = [.242, .552], pd = 100%, p = .0, ps = 100%, pl = 89.23%), such that the more similar a trait is to the ingroup and/or dissimilar from the outgroup, the more self-descriptive it is. However, we find no evidence for the metacontrast ratio's association with self-evaluations differing for the lower status (β = -.223, OR = .799, $CI_{95\%}$ = [-.626, .191], pd = 86.80%, p = .2640, ps = 80.88%, pl = 35.18%) or higher status (β = -.132, OR = .876, $CI_{95\%}$ = [-503, .246], pd = 75.78%, p = .4845, ps = 100%, pl = 89.23%) conditions (Figure 11A). Thus, the Similarity-to-ingroup relative to Similarity-to-outgroup is associated with self-evaluations, but this does not appear to depend on the intergroup contrast in this university context.

We further interrogate the nature of these effects by separating summed similarity into distinct regressors. Consistent with the absence of evidence for outgroup differentiation effects in the context of self-concept generalization, we found that the summed similarity to ingroup trait classifications (β = .493, OR = 1.638, $CI_{95\%}$ = [.367,

.630], pd = 100%, p = .0, ps = 100%, pl = 99.99%), but not the summed similarity to outgroup trait classifications (β = .062, OR = 1.064, $CI_{95\%}$ = [-.082, .201], pd = 80.98%, p = .3805, ps = 56.58%, pl = 0.05%), was associated with the self-descriptiveness of traits (Figuree 11B, Figure 11C). Interestingly and perhaps surprisingly, Similarity-to-outgroup was most negatively associated with self-evaluations for the Negation comparison, which may suggest that people are more repelled by what is not their ingroup rather than what is an outgroup university. Again, this suggests that it may be moreso that it is a motivation to see oneself as similar to the ingroup than a motivation to see oneself as different from the outgroup than accounts for self-anchoring processes.

Figure 11. The association of self-evaluations with Metacontrast Ratio, Similarity-to-ingroup, and Similarity-to-outgroup.



Note. Plotted using marginal effects with 1.96 +/- SEs . Top row is Study 2; Bottom row is Study 3. Left column is Metacontrast Ratio as predictor, middle column is Similarity-to-ingroup as predictor, right column is Similarity-to-outgroup as predictor. Y-axis are Likert self-evaluations. X-axis predictors are Z-scored. For the purpose of visualization and ease of interpretation, predictions are treated as continuous variables and graphs are depicted on the latent Likert scale rather than plotting the outcome on the ordinal scale which would require seven panels per plot.

Correlates and Differences in Trait Segregation

We perform further exploration of the individual differences that are associated with the segregation of traits along group boundaries. We find that individuals higher in need to belong ($r = .13$, $CI_{95\%} = [0.01, 0.24]$) and dialectical self-views ($r = .18$, $CI_{95\%} = [0.05, 0.28]$) segregate traits more based on groups, whereas individuals higher in self-esteem ($r = -.17$, $CI_{95\%} = [-0.28, -0.05]$), independence ($r = -.23$, $CI_{95\%} = [-0.34, -0.11]$), and social identification ($r = -.15$, $CI_{95\%} = [-0.26, -0.04]$) segregate traits less based on groups. Notably, the tendency to segregate traits more for individuals higher in need to belong and to segregate traits less for more independent and higher self-esteem individuals are replications of Study 1. Again, these findings may provide insights into the types of individuals more prone to carving distinctions between groups regardless of self-beliefs.

We additionally find that individuals differed in their tendencies to segregate trait classifications based on what outgroup they were comparing the ingroup against. Specifically, those who compared their ingroup against the higher status university ($M = .11$, $SD = .08$) segregated traits significantly more than the lower status ($\beta = -.754$, $CI_{95\%} = [-1.181, -.652]$, $pd = 100\%$, $p = .0$, $ps = 100\%$, $pl = 100\%$; $M = .05$, $SD = .07$) and negation ($\beta = -.912$, $CI_{95\%} = [-1.017, -.488]$, $pd = 100\%$, $p = .0$, $ps = 100\%$, $pl = 100\%$; $M = .04$, $SD = .05$) comparison groups. Thus, these individuals that compared their university group to a higher status group appeared to rely on more strict differentiation of the ingroup and outgroup, rather than self-projection per se. Part of this may be evidenced by the fact that individuals in the higher status condition exhibited significantly lower

intercepts overall, and thus a persistently lower likelihood of classifying traits as characteristic of the ingroup in general. This may be due to the fact that the sample consists of largely positive traits, and people may be classifying more traits as characteristic of the higher status outgroup in general, regardless of self-beliefs.

Generalization Model

We next attempted to replicate the Generalization Model previously established in Study 1. We fit the computational models separately for each condition.

Model Performance. The self-projection model converged for both the majority and minority comparison conditions: there were no Rhat values above 1.01 for the "Not UCR" condition, there were 31 Rhat values above 1.01 for the Higher Status condition, and 85 Rhat values above 1.01 for the Lower Status condition. However, none of these three were below the more liberal Rhat threshold of 1.05 (or even 1.023). Across all three conditions, the self-projection model outperformed the simpler bias-only model.

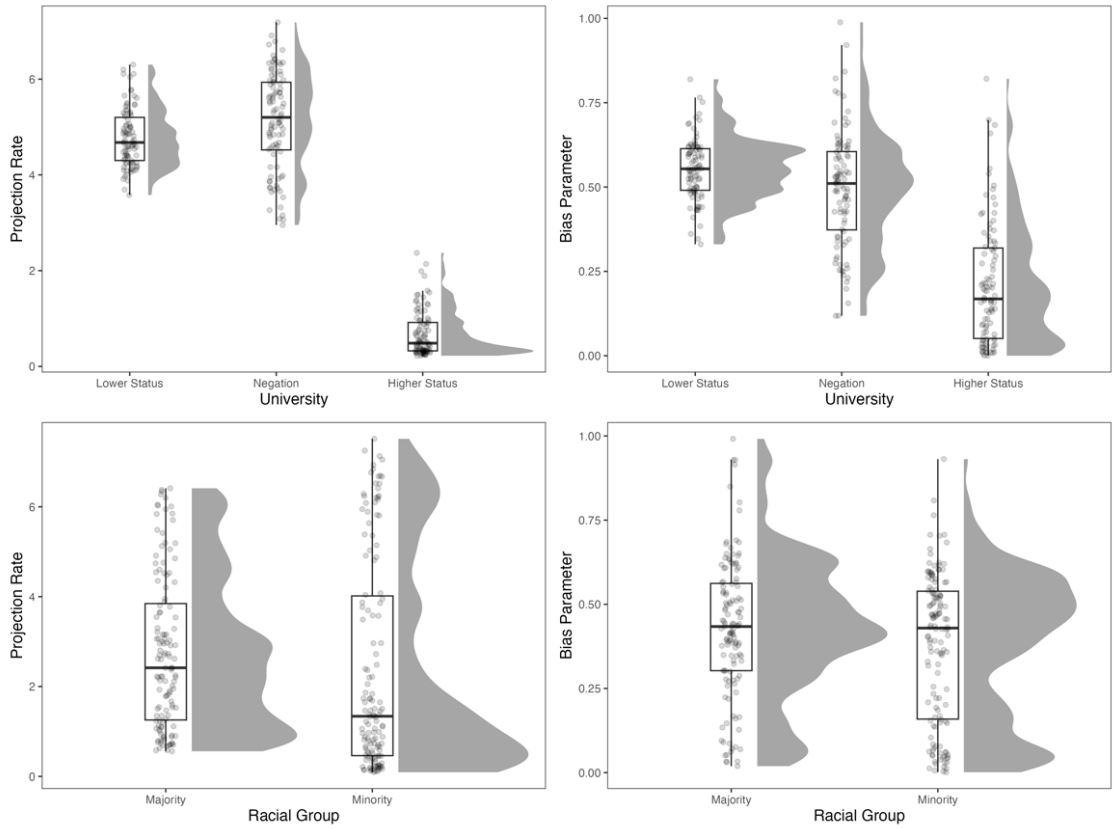
Parameter Recovery. We examined parameter recovery using the original fitted parameters estimated from each condition. For the higher status condition, the temperature ($r = .72$), projection rate ($r = .85$), and bias parameter ($r = .63$) were all recoverable. For the lower status condition, the temperature parameter ($r = .76$) and bias parameter ($r = .70$) were recoverable, and the projection rate was modestly recoverable ($r = .52$). For the negation condition, the temperature parameter ($r = .86$) and the bias parameter ($r = .78$) were recoverable while the projection rate was not very recoverable ($r = .37$). Overall, despite weak recoverability in one condition, we believe that the recoverability of the parameters overall is supported.

Differences in Parameters. We were additionally interested in whether there would be differences in the *projection rate* by condition. We found that participants who compared their university against a higher status comparison (M = .68, SD = .48) exhibited lower projection rates than both those who compared against the negation comparison ($\beta = 2.053$, $CI_{95\%} = [1.957, 2.149]$, pd = 100%, p = .0, ps = 100%, pl = 100%; M = 5.13, SD = 1.02) and the lower status comparison ($\beta = 1.887$, $CI_{95\%} = [1.789, 1.987]$, pd = 100%, p = .0, ps = 100%, pl = 100%; M = 4.77, SD = 0.60). Findings suggest that individuals may be less extreme in their tendencies to convert self-beliefs to ingroup-beliefs, that are then generalized based on similarity, when compared against a higher status outgroup (see Figure 12). Given that the higher status outgroup is more positively perceived, it may be that people are less extreme in how they convert their self-beliefs into ingroup beliefs when they feel more favorable towards to the outgroup and they are motivated to see their own attributes in the higher status outgroup.

Correlations in Parameters. We next examined the correlations among the projection rate and individual differences measures. Collapsed across all conditions, we found that the projection rate was correlated with social identification (r = .20, $CI_{95\%} = [0.09, 0.31]$), intergroup bias (r = .18, $CI_{95\%} = [0.05, 0.31]$), and Private Collective Self-Esteem (r = .17, $CI_{95\%} = [0.06, 0.28]$). We thus replicate that social identification is correlated with projection rate, and provide an extension in the current study involving real groups that it is also associated with intergroup bias (i.e., liking of one's ingroup university and disliking of outgroup universities). Thus, the extent to which people convert self-beliefs into ingroup beliefs is reflected in people's ingroup attachments.

Corroborating the interpretation that the lower projection rates observed in the higher status outgroup comparison condition relative to the other two conditions are due to feeling more positively towards the prestigious university and being motivated to see oneself as similar to the higher status outgroup university, for the higher status condition, a lower projection rate was associated with self-outgroup overlap with the higher status outgroup ($r = -18$, CI = [-.36, .01]) and ingroup-outgroup overlap with the higher status outgroup ($r = -.16$, CI = [-.34, .03]). Thus, generally the projection rate is associated with attachment to one's own group and disliking of other groups, but when contrasting one's ingroup against the higher status outgroup, people self-project less to the extent that they perceive themselves or their ingroup as more similar to the higher status outgroup.

Figure 12. Raincloud plot depicting differences in projection rate and bias parameter across conditions.



Note. Top row is Study 2; Bottom row is Study 2. Y-axis depicts the parameter, while X-axis depicts the condition.

Discussion

In this study, we replicate the previous findings that people generalize from the self-concept to the ingroup on the basis of similarity to the self. In terms of outgroup comparisons, we find evidence that people may be less extreme in converting self-beliefs into ingroup-beliefs when the outgroup comparison is the higher status outgroup. Although the focus was on the status of the outgroup, this outgroup university was also more positively perceived than the lower status outgroup while the negation comparison had no explicit affective attributions attached to it. Thus, it may be that when the outgroup comparison is a more positively perceived and higher status outgroup, that people may feel less motivation to project their self-beliefs onto the ingroup, but rather may instead be motivated to project self-beliefs more onto the outgroup as well in order to perceive themselves as more like the "prestigious" outgroup university. We additionally find that the projection rate is associated with individual differences in intergroup bias and social identification, reflecting that people who dislike outgroups more and like their ingroup more may be more motivated to project themselves onto the ingroup. Alternatively, individuals who project themselves more onto the ingroup and less onto the outgroup may experience greater ingroup bias as a result. The causal relationship between intergroup bias and self-projection should be further elucidated in future work, potentially incorporating an intergroup cooperation or bias behavioral task with a self-anchoring task such as this.

Study 3: Self-Anchoring Based on Relative Size of Racial Groups

Members of minority groups often exhibit stronger social identification because they are relatively smaller in size, which allows people to experience greater perceived distinctiveness from other larger social groups (Leonardelli et al., 2010). Minority groups are also perceived as more homogenous (Simon & Brown, 1987), more entitative (Mullen, 1991), and more similar to each other (Nelson & Miller, 1995). Therefore, to the extent that two minority groups are contrasted against one another, this may amplify distinctiveness between the groups and increase the motivation to self-anchor to the ingroup. Alternatively, minority-majority intergroup differences often signify status differences (Fiske et al., 2016), and contrasting one's minority ingroup against a majority outgroup may evoke beliefs about status-based differences and lead to a desire to be less like the lower status group. In the current study, we extend self-anchoring to be applied to racial groups for the first time, while also examining whether majority or minority outgroup contrasts exert differential effects on the tendency to self-anchor and generalize the self-concept to the ingroup.

Methods

Participants

We recruited 313 participants and excluded 48 participants to arrive at a final sample of N = 265. We excluded any participants who self-reported not taking the task seriously (less than '4' to "To what extent did you take this task seriously?"; N = 2), who self-reported that their data is unusable ("No" to "Did you understand the task and respond truthfully and meaningfully enough that your data is usable?"; N = 9). We

additionally excluded participants if they exhibited behavior during the task that reflected careless responding, with exclusionary criteria including if over 80% of their self-evaluations were identical, if over 95% of their group classification were identical, or if over 40% of their behavioral responses were missing from either part of the task (N = 37).

Participants (N = 265) were native English-speaking university students (59.14% Cisgender Female, 40.86% Cisgender Male; M_{Age} = 19.01, SD_{Age} = 1.71, $Range_{Age}$ = [17, 33]), and were 50.97% Hispanic/Latino and 49.03% Asian. We conducted a summary statistics based power analysis for mixed-effects models (Murayama et al., 2022) from Study 2 using data from before study was complete with a sample of 180. A cross-level interaction between similarity-to-self for each trait interacting with condition (t = 2.547) requires a sample of N = 220 for 80% power.

Design

The task was similar to Study 2, except that rather than being asked to classify traits as characteristic of a university ingroup, participants were asked to classify traits as characteristic of a racial ingroup or outgroup. Specifically, exclusively Asian or Latino participants were recruited. Participants were assigned to one of two conditions: Minority or Majority comparison. If a Latino or Asian participant was assigned to the Majority outgroup condition, they compared their racial identity to "White". Meanwhile, if a Latino participant was assigned to the Minority outgroup condition, they compared their racial identity to "Asian", while if an Asian participant was assigned to the Minority outgroup condition, they compared their racial identity to "Latino".

Measures

The same measures were used as in Study 2. The identity-specific questions were adapted to pertain to participants' Latino or Asian racial identity. We attempted to collect the self-group overlap scale again but there were data collection issues and the data from this inventory was unusable in Study 3.

Planned Analyses

We perform identical analyses as in Study 2, except using the two-level condition factor as a dummy coded (Majority is reference group) moderator of effects. We conducted additional analyses controlling for the racial identity of the participant to see if the effects differed for Latino and Asian participants.

Results

Differences Across Racial Groups in Perceived Warmth and Status

We again examined how participants perceived the social status and their warmth towards each of the social groups involved in the study. We found that compared to the White social group (M = 8.42, SD = 1.63), Latino (β = -4.322, OR = .013, $CI_{95\%}$ = [-4.760, -3.872], pd = 100%, p = .0, ps = 100%, pl = 100%; M = 5.02, SD = 1.59) and Asian (β = -2.194, OR = .111, $CI_{95\%}$ = [-2.553, -1.830], pd = 100%, p = .0, ps = 100%, pl = 100%; M = 6.73, SD = 1.51) groups were perceived as lower in social status, while Latino were perceived as lowest in social status.

We next examined whether there were differences in the perceptions of warmth towards each group. We found that compared to the White social group (M = .54, SD = .23), Latino (β = .974, $CI_{95\%}$ = [.823, 1.128], pd = 100%, p = .0, ps = 100%, pl = 100%;

M = .79, SD = .19) and Asian (β = .860, $CI_{95\%}$ = [.707, 1.011], pd = 100%, p = .0, ps = 100%, pl = 100%; M = .77, SD = .19) groups were perceived more positively. Across the sample consisting of Latino and Asian participants, the White outgroup was not perceived positively (see Figure 10). This draws an interesting contrast with Study 2, where the higher status outgroup was perceived relatively positively (almost as positively as the ingroup), while in the current study, the higher status outgroup is perceived quite negatively relative to the focal two minority ingroups.

People Classify Desirable Traits as Ingroup Characteristic

We again test whether people are motivated to differentially attribute more positive traits to the ingroup than the outgroup. We again find that people are more likely to classify desirable traits as characteristic of their racial minority ingroup than the outgroup, but that this effect is not likely to be large (β = .163, OR = 1.178, $CI_{95\%}$ = [.086, .239], pd = 100%, p = .0, ps = 96.93%, pl = 0%). We further interrogate whether this effect differs depending on the outgroup comparison and find that when people compared their racial minority ingroup against a racial majority outgroup (i.e., White), the tendency to project desirable traits to the ingroup was stronger, although this difference in slopes was not likely to be large (β = -.106, OR = .899, $CI_{95\%}$ = [-.227, .106], pd = 95.50%, p = .0, ps = 59.95%, pl = 0%). This may reflect a stronger motivation to establish positive distinctiveness from an outgroup via ingroup enhancement if the outgroup is the superordinate and higher status majority group (Figure 7G).

People Classify Self-Descriptive Traits as Ingroup Characteristic

We next examined the extent to which people self-anchor on repeated traits based on the correspondence between trait self-evaluations and later ingroup classifications. Again, we found that people were highly likely to characterize self-descriptive traits as characteristic of the ingroup (β = .208, OR = 1.232, $CI_{95\%}$ = [.153, .265], pd = 100%, p = .0, ps = 100%, pl = 0%). We next implemented an interaction by condition and found that there was weak evidence that people self-anchor more based on repeated self-evaluations when the majority is the comparison (β = -.010, OR = .905, $CI_{95\%}$ = [-.212, .016], pd = 95.25%, p = .0950, ps = 55.90%, pl = 0%). This again may reflect that people are more likely to self-project in a manner that distinguishes from the outgroup if the outgroup is the higher status and more disliked majority outgroup (Figure 7H).

People Classify Traits Similar to Self-Descriptive Traits as Ingroup Characteristic

While the previous analysis demonstrates that people may project self-characteristics to the racial ingroup on repeated traits, especially when compared against a majority outgroup, we attempted to examine whether this effect is reflected in semantic generalization across all traits, not just repeated traits. We found that similarity-to-self predicts ingroup classifications, regardless of condition (β = .692, OR = 1.997, $CI_{95\%}$ = [.290, 1.068], pd = 100%, p = .0, ps = 99.95%, pl = 76.43%). Once again, we find that this more mechanistic approach of modeling self-anchoring is a much stronger predictor of ingroup classifications than mere repeated self-evaluations (Figure 7I).

We next estimate an interaction and found that there was no evidence that Similarity-to-self differed for the minority condition relative to majority condition (β =

.331, OR = 1.997, $CI_{95\%}$ = [-1.106, .414], pd = 80.85%, p = .3830, ps = 73.25%, pl = 29.03%). Interestingly, while people's tendencies to self-project to the ingroup on repeated traits based on self-evaluations appears to be stronger when the outgroup is majority than minority, this difference may not be reflected when utilizing similarity-based generalization.

We next examined whether this tendency to generalize using similarity-to-self is equivalently robust across both repeated and novel traits. To do so, we estimate the interaction between similarity-to-self and novelty (as a dummy coded categorical factor). We conducted an equivalence test and found that the differences between the slopes for novel and repeated traits were approximately equivalent to zero. We found that the difference between the similarity-to-self slopes for novel and repeated traits was indeed equivalent to zero (ROPE = [-.18, .18], $CI_{95\%}$ = [-.06, .03], inside ROPE = 100%), suggesting that generalization is approximately equivalent regardless of whether traits were previously self-evaluated or not. We next explored whether this tendency to generalize across novel and repeated traits using similarity-to-self differs between conditions, by estimating a three-way interaction between novelty, condition, and similarity-to-self. We found insufficient evidence to suggest that the generalization of similarity-to-self to repeated and novel traits differs between conditions (β = .072, OR = 1.075, $CI_{95\%}$ = [-.023, .168], pd = 93.50%, p = .0, ps = 36.20%, pl = 0%). Interestingly, there is not much evidence to suggest that the effect of similarity-to-self differs between outgroup comparisons either.

Finally, we tested whether the effect of similarity-to-self on ingroup classifications depends on outdegree or indegree centrality. We found insufficient evidence that the effect of similarity-to-self on ingroup classifications depends on outdegree (β = -.002, OR = .997, $CI_{95\%}$ = [-.034, .031], pd = 56.18%, p = .8765, ps = 0%, pl = 0%). However, we found some weak evidence that the effect of similarity-to-self on later ingroup classifications is stronger for higher relative to lower indegree traits (β = -.019, OR = .980, $CI_{95\%}$ = [-.044, .004], pd = 94.80%, p = .1040, ps = 0%, pl = 0%). This effect is small but across the three studies, there is suggestive evidence that traits with inputs are more strongly generalized to using similarity. This may be because these traits receive more information from other traits, and thus are more readily generalized to.

Self-Uncertainty Predicts Less Ingroup Classification

In Studies 1 and 2, we found that traits are less likely to be classified as characteristic of the ingroup if they are more self-evaluatively uncertainty. We explored this again in Study 3 and found insufficient evidence to support the same effect (*β* = -.187, *OR* = .829, $CI_{95\%}$ = [-.488, .116], *pd* = 87.93%, *p* = .2415, *ps* = 73.08%, *pl* = 0.98%), although this effect was directionally the similar but with larger error around the estimate. Although this effect is inconsistent with Study 1 and 2, the similar direction may reflect that consistent findings with a large amount of noisiness around this effect.

Projection to the Ingroup, Rather than Rejection of the Outgroup

The previous analyses suggest that a trait's similarity-to-self predicts how likely it is to be perceived and classified as ingroup characteristic. We next examined whether a trait's metacontrast ratio is associated with its self-descriptiveness. We found that traits

with a higher metacontrast ratio were evaluated more self-descriptively (β = .194, OR = 1.215, $CI_{95\%}$ = [.083, .306], pd = 99.95%, p = .001, ps = 99.25%, pl = 3.35%). Again, this suggests that to the extent that trait's are more similar to the ingroup and less similar to the outgroup, they are also evaluated more self-descriptively. We further examine whether the association between the metacontrast ratio and self-evaluations differs between outgroup comparisons. We find that the association between the metacontrast ratio and self-evaluations is stronger when the outgroup comparison is a Majority groip than when the outgroup comparison is a Minority group (β = -.203, OR = .816, $CI_{95\%}$ = [-.430, .019], pd = 96.33%, p = .0735, ps = 90.75%, pl = 20.23%). This may reflect that people establish greater distinctiveness from the outgroup by self-projecting more onto the ingroup when the majority is the comparison, as minority members may be more motivated to avoid seeing themselves in the more disliked majority group than the more positively perceived minority outgroup (Figure 11D).

We split the metacontrast ratio into separate regressors of summed similarity to ingroup or outgroup, and predicted self-evaluations of traits during training phase. Again, we find that a trait's similarity-to-ingroup classified traits (β = 1.361, OR = .816, $CI_{95\%}$ = [.189, .423], pd =100%, p = .0735, ps = 100%, pl = 55.55%), but not outgroup classified traits (β = .051, OR = 1.053, $CI_{95\%}$ = [-.070, .163], pd = 80.68%, p = .3865, ps = 50.98%, pl = 0%) is associated with its self-descriptiveness. We additionally investigate whether this association between similarity-to-ingroup or similarity-to-outgroup and self-evaluations differs between outgroup comparisons. We found that similarity-to-outgroup is more negatively associated with self-evaluations when the majority is the outgroup

compared to the minority (β = .142, OR = .935, $CI_{95\%}$ = [-.023, .311], pd = 95.28%, p = .0945, ps = 85.28%, pl = 3.43%). There was insufficient evidence to suggest that similarity-to-ingroup's association with self-evaluations differed by outgroup comparison (β = -.067, OR = .935, $CI_{95\%}$ = [-.215, .082], pd = 81.58%, p = .3685, ps = 59.20%, pl = 0.05%). This may suggest that while Similarity-to-outgroup is not generally associated with self-evaluations, there is greater tendency to reject the outgroup in order to differentiate when the outgroup is the higher status and more disliked majority (Figure 11E and 11F). The motivation to repel from the majority outgroup may be thus driving what is contributing to the differences in the effect of metacontrast ratio on self-evaluations. Furthermore, supporting the interpretation that during the minority outgroup contrast, people may see aspects of themselves more in other minority outgroup members than majority outgroup members, Similarity-to-outgroup is positively associated with self-evaluations in the minority condition but not the majority condition.

Correlates and Differences in Trait Segregation

Again, we examined whether the extent to which participants segregated traits in the network based on group classifications correlated with various individual differences. Interestingly, participants who perceived White people as being higher in status than Asian (r = .17, CI = [.05, .29]) and Latino (r = .22, CI = [.10, .33]) people segregated traits between groups more. Again, we replicated trait segregation's association with need to belong (r = .18, CI = [.06, .29]). In contrast, individuals who were more strongly identified (r = -.15, CI = [-.27, -.04]), whose racial identity was more central (r = -.17, CI = [-.29, -.05]), who had more positive feelings towards their racial identity (r = -.13, CI =

[-.25, -.01]), and who perceived their own racial group members as more similar to each other (r = -.12, CI = [-.24, -.01]) segregated traits less between groups. Thus, in the racial context, it appears that individuals who are more strongly identified and feel more positively about their social group segregate traits less based on group classifications, whereas individuals who perceive greater status differences between the White majority and the racial minority groups segregate traits more. It may be that trait segregation reflects a tendency to persist in differentiating groups regardless of self-beliefs, which is greatest in individuals who perceive larger status differences between minority and majority groups. Meanwhile, those who strongly identify with or feel positively about their ingroup racial identity may segregate traits less as they are classifying group attributes based on self-beliefs, and not merely distinguishing groups in a coarse manner.

In a separate test, we examine whether there are differences in the tendency to segregate traits based on group classifications depending on the outgroup contrast. We find that relative to individuals who compare their racial ingroup against a majority outgroup (M = .08, SD = .07), individuals who compare their racial ingroup against a minority outgroup segregate traits more (β = .522, $CI_{95\%}$ = [.293, .764], pd = 100%, p = .0, ps = 100%, pl = 100%; M = .12, SD = .08). As in Study 2, it appears that the comparison for which less self-projection to the ingroup occurs also has greater Trait Segregation. This may reflect that when people self-project less to the ingroup, they classify group characteristics based on coarser distinctions.

Generalization Model

Model Performance. The self-projection model converged for both the majority and minority comparison conditions: there were no Rhat values above 1.01 for the majority comparison condition, while there were only three Rhat values above 1.01 for the minority condition and none of these three were below the more liberal Rhat threshold of 1.05. For both conditions, the self-projection model outperformed the null bias-only model.

Parameter Recovery. Using the original parameters to simulate behavior, for the majority condition, we found that the temperature (r = .70) and bias parameters (r = .66) were recoverable, while the projection rate was only modestly recoverable (r = .44). For the minority condition, we found that the temperature (r = .46), projection rate (r = .55), and bias parameter (r = .57) were modestly recoverable.

Differences in Parameters. We estimated the difference in the projection rate for the majority relative to the minority condition, while controlling for the racial identity of the participant. We found that the projection rates for the majority condition (M = 2.76, SD = 1.7) were higher than the projection rates for the minority condition (β = -.237, $CI_{95\%}$ = [-.493, .007], pd = 97.18%, p = .0565, ps = 85.50%, pl = 0.1%; M = 2.38, SD = 2.35). This may reflect that people self-project more onto the ingroup when comparing against the more disliked majority outgroup. We also found that Latino participants (M = 2.92, SD = 2.02) on average had higher projection rates than Asian participants (β = .386, $CI_{95\%}$ = [.142, .630], pd = 99.95%, p = .0010, ps = 85.50%, pl = 0.1%; ; M = 2.19, SD = 2.06). Our prior analyses showed that while the majority outgroup is perceived as higher

status, they are also perceived less positively overall. It may be that people project less when contrasting against the majority group because they experience more dislike towards them, and are thus more motivated to differentiate themselves from the more disliked outgroup (Figure 12).

Correlations in Parameters. We next explored the associations of the projection rate across the full sample. Among the strongest positive associations were social identification (r = .11, CI = [-.01, .23]), independence (r = .10, CI = [-.02, .22]), and need for cognition (r = .12, CI = [0, .24]). Meanwhile, dialectical self-views were negatively correlated with projection rate (r = -.21, CI = [-.32, -.09]). The association is weaker than in Study 2, but again we replicate that people who more strongly identify with their social group also exhibit higher projection rates. However, we were surprised not to replicate that individual differences in intergroup bias were associated with projection rates.

In terms of the bias parameter, people are more biased towards outgroup classifications to the extent that they are more independent (r = -.19, CI = [-.30, -.07]), higher in self-esteem (r = -.19, CI = [-.30, -.06]), need for cognition (r = -.18, CI = [-0.30, -0.07]), self-concept clarity (r = -.17, CI = [-0.27, -0.04]). We replicate that individuals who are more independent exhibit higher bias towards outgroup classifications, reflecting that such individuals may be more prone to perceiving and classifying traits as outgroup characteristic, regardless of self-beliefs.

Discussion

One potential explanation for the fact that people exhibit greater projection rates when contrasting their minority racial identity against a majority racial identity (e.g.,

White) may be that people are more motivated to engage in differentiation (R. Brown, 2000; R. J. Brown, 1984) under this comparison. Minority members may perceive "common fate" (D. T. Campbell, 1958; Sell & Love, 2009) with other minority members, due to shared experiences of marginalization or discrimination as minority members, and thus self-project less onto their minority ingroup exclusively. Additionally, it is important to note that the White majority is perceived much less positively than either minority group, which may be contributing to the differences, given that the projection rate is also associated intergroup bias across Study 2 and Study 3. Differences in perceptions of positivity towards each group driving the effect would align with the findings from Study 2, whereby the projection rate was lower when contrasted against the higher status but more positively perceived outgroup.

General Discussion

People self-project their own self-perceived attributes onto similar others (Ames, 2004) or onto ingroup members (M. R. Cadinu & Rothbart, 1996), but little is known about the within-person cognitive mechanisms underlying this process, such as the mechanisms by which people engage in these inferences and what contextual factors influence these inferences. We rely on a semantic, feature-based model of similarity to examine how people infer their ingroup is similar to themselves, and find that people engage in similarity-based generalization, classifying that their ingroup is like them on related, but different, traits. Further, we find that people are less extreme in their tendencies to convert self-beliefs into ingroups beliefs if the outgroup is a higher status but more liked outgroup, or if the outgroup is a fellow racial minority outgroup that is

more liked. Finally, we find that people's tendency to project to the ingroup may be primarily driven by a motivation to achieve similarity with the ingroup, rather than a motivation to achieve differentiation with the outgroup.

Formalizing Social Identity Theory

The Social Identity Approach (Tajfel, 1978; Turner et al., 1987) has borrowed heavily from cognitive science research on concept learning and representation, describing an individual's prototypicality of the ingroup by the metacontrast ratio between their similarity to the ingroup over their similarity to the outgroup (D. T. Campbell, 1958; Rosch & Lloyd, 1978). However, Social Identity Theory has rarely implemented the formalism leveraged by the concept learning research and theory that it draws inspiration from. The ability to make representational claims about the role of similarity-based inference in social categorization and group-based inference is limited by the absence of model formalism (Guest & Martin, 2021; Robinaugh et al., 2021). A reason for the preponderance of verbal descriptions of similarity but little to no formal measurements of similarity in previous social identity research may be that it is difficult to operationalize formal measures of similarity of oneself to the ingroup or the outgroup. Instead, intragroup process and social identification research often rely on self-report inventories assessing individual differences in perceived similarity to the ingroup (Fielding & Hogg, 1997; Hogg & Hains, 1996). Meanwhile, most research on category learning has primarily focused on stimuli belonging to perceptual categories, where stimuli can be mathematically separated in psychological distance by embedding them in multidimensional Euclidean space (Shepard, 1987). Using a feature-based model of

semantic similarity (Tversky, 1977), whereby traits are more similar if they share more semantically related connections (i.e., neighbors) in common, the current studies bridge the model formalism of past category learning research with the allusions to category learning theory expressed in the Social Identity Approach.

Self-Anchoring as Generalization Across Related Traits

Using this approach, we flexibly estimate measures denoting Similarity-to-self, Similarity-to-ingroup, and Similarity-to-outgroup, which provide novel insights that were not previously established in prior research on self-anchoring. For example, prior research on self-anchoring suggests that it is an inductive reasoning process (DiDonato et al., 2011; Krueger, 2007) implicating generalization (reasoning about group members based on an N = 1 of the self-concept), but has entirely relied on comparing evaluations on repeated traits (van Veelen, Otten, et al., 2016). This prior research could thus partially be interpreted as due to repetition effects, whereby ratings on the same items on repeated occasions amplify the descriptiveness of the items (Unkelbach & Rom, 2017). The current research strengthens the generalization claims by finding that participants generalize from the self to the ingroup across related traits, even for novel traits that were not previously self-evaluated. Thus, people infer what their group ought to be on related traits, based on how they self-evaluated previously and the similarity to the related traits (i.e., if I am sociable, my group ought to be fun and witty, but not necessarily disciplined). As such, the current findings provide stronger evidence of generalization, reflecting that people not only generalize from themselves to their group members, but also generalize across related attributes and beliefs when doing so.

We additionally find that trait self-evaluations are primarily associated with Similarity-to-ingroup, and not as much Similarity-to-outgroup. In this specific context when undergoing generalization from the self to the ingroup, people may be more motivated by ingroup love over outgroup hate (Brewer, 1999; Lelkes & Westwood, 2017), preferring to self-project onto the ingroup than to distinguish themselves from the outgroup (Gramzow et al., 2001; Otten, 2003). However, contextual effects such as competition and threat can alter the degree to which people express ingroup favoritism or outgroup punishment and the tendency to self-project onto the ingroup may vary under contexts where greater animosity, conflict, and intergroup tension is present. This is supported by the fact that Similarity-to-outgroup was more negatively associated with self-evaluations when the Majority outgroup was the comparison, reflecting potentially greater intergroup tension towards the more disliked outgroup. Thus, in such contexts, people may be motivated to see themselves as unlike the outgroup as much as like their ingroup.

Contrastive Effects Augment Self-Projection

Here, we find that people attenuate how strongly self-beliefs are converted to ingroup beliefs when comparing their ingroup against a higher status but more warmly regarded outgroup university, and also when comparing against a warmly regarded but lower status fellow racial minority outgroup. This dovetails with the Generalized Context Model (Nosofsky, 1986, 2011), which suggests that people categorize stimuli into categories on their basis of similarity to existing exemplars of a category, and this likelihood decays exponentially as a function of dissimilarity from the target stimulus

(i.e., a 'robin' might be classified as a 'bird' based on its similarity to other exemplars of bird category such as 'sparrow' and 'pigeon'). However, the context in which classifications are made augment the psychological structure within which stimuli are embedded (i.e., humans and mannequins may be judged as highly similar in a context that emphasizes structure but dissimilar in a context that emphasizes vitality). Together, this may reflect that people may shrink the self-descriptiveness weights applied to semantic similarities in contexts where the outgroup is more warmly regarded and there is less intergroup animosity. This aligns with Social Identity Theory which suggests that people engage in social identification and alignment with the ingroup to achieve positive differentiation from the outgroup (Tajfel, 1974, 1978) but to the extent that there are not feelings of animosity or conflict, there is less need to positively differentiate from the outgroup. Diminished projection rates may thus index greater perceived harmony with a contrasted outgroup, such that individuals are able to see aspects of themselves in both the ingroup and outgroup.

Similarity with the Ingroup as a Pathway to Social Identification

Social identification with one's social groups has important implications for behavior, including ingroup favoritism (Balliet et al., 2014; Hewstone et al., 2002) and intergroup bias (Obst et al., 2011). One possible route to social identification is self-anchoring (van Veelen, Otten, et al., 2016), and one of the original motivations for investigating self-anchoring as a mechanism for social identification, was the fact that mere assimilation of group attributes into the self-concept (i.e., self-stereotyping) alone cannot explain the tendency to engage in ingroup favoritism for groups which individuals

have no prior knowledge about, such as minimal groups (M. R. Cadinu & Rothbart, 1996; van Veelen et al., 2013a). In the current study, a computational parameter reflecting the extent to which individuals project to the ingroup– and reject the outgroup– correlated with individual differences in social identification and intergroup bias. Individuals who are more extreme in their tendencies to self-project to the ingroup and reject the outgroup in their similarity-based generalization across traits may be more likely to be strongly identified with their ingroup and feel that they are representative of their ingroup. Alternatively, strongly identified individuals may be more likely to project their own self-beliefs onto the ingroup in order to maximize assimilation with the ingroup. Additionally, the projection rate is associated with intergroup bias in the university context, corroborating that intergroup bias may be manifested by the extent to which individuals perceive themselves as similar to their ingroup and dissimilar from outgroups (DiDonato et al., 2011; Roth et al., 2018; Schubert & Otten, 2002). Future research should investigate the extent to which these individual differences in projection-to-ingroup and rejection-of-outgroup may translate to greater ingroup bias.

Generalization for Superordinate or Subordinate Group Identities

People are often more strongly identified with and feel more meaningfully attached to smaller, more inclusive and differentiated subordinate groups (i.e., subgroups) than larger, superordinate groups (Brewer, 1991; Leonardelli et al., 2010). In the tradition of Social Identity Theory (Tajfel, 1978), Self-Categorization Theory (Turner et al., 1987) and the broader Social Identity Approach, social identities are not aspects within an individual's self-concept, but rather are extensions of the self. Social identities are

categorizations of the self into social units, and the self-concept is depersonalized when conceptualizing oneself in terms of these higher levels of abstraction. For example, when considering oneself as an individual, one may consider the beliefs and characteristics that distinguish oneself from others. Meanwhile, while representing oneself as a graduate student in the psychology department at UCR, the salient features of one's self-concept are the various features that one shares in common with other UCR students, while when representing oneself as an academic more broadly, the salient features of one's self-concept may be the various features that one shares in common with other academics more generally (Brewer, 1991). People often identify with meaningful subgroups over superordinate groups (Brewer, 1993; Hornsey & Hogg, 1999), contribute more to shared resources when subgroup rather than superordinate group is relevant (Rabinovich & Morton, 2011), and express greater intergroup bias when a subgroup identity is relevant (Hornsey & Hogg, 2000). In contrast, thinking of oneself as a member of a common, superordinate group identity is associated with greater social harmony and cohesion (Gaertner et al., 2016). Social Identity Theory's perspective on self-definition varying at different levels of abstraction focuses primarily on how "I" shifts to "We" (i.e., self-stereotyping). However, rarely discussed is how "We" shifts to "I" (i.e., self-anchoring) at varying levels of abstraction. Indeed, there is some evidence that self-stereotyping contributes to social identification and well-being for minority groups but not majority groups (Latrofa et al., 2009, 2012), while there is less research on how self-anchoring may be involved in social inferences for subordinate versus superordinate social groups. There is some evidence that because individuals of minority subgroups feel less similar to

the superordinate group, self-anchoring is a more effective means to achieve social identification with superordinate categories for members of subgroups (van Veelen et al., 2013b). Future work may contrast superordinate and subordinate ingroup categories to determine which is more likely to be generalized to, and how this influences social identification towards one level of abstraction over another.

Limitations and Extensions

Some aspects of the current work could be improved on. Namely, the current research was conducted on undergraduate students, and self-anchoring is considered a possible mechanism for group identification not only for newcomers to groups, but also for individuals early in development as they acquire knowledge about social groups (van Veelen, Eisenbeiss, et al., 2016). Personality and behavior varies across the lifespan (G. H. Elder, 1998), and future research may attempt to further extend this concept generalization framework to examine how the tendency to generalize from the self to the group differs across age ranges. It may be that adolescents are more likely to self-project than older adults, due to having less established beliefs and knowledge about existing groups.

In terms of design, estimating separate computational parameters for the ingroup projection and outgroup rejection was not possible due to issues with parameter identifiability (Guillaume et al., 2019); there was not a sufficient number of observed variables to distinguish separate projection rates for outgroup and ingroup choices. In order to identify separate parameters for each ingroup and outgroup projection rate, future research may consider producing more types of self-beliefs or features from which

to generalize. Greater dimensionality in the design will allow future researchers to identify separate parameters for outgroup repulsion and ingroup projection. Additionally, the current research used entirely positive traits. Ideally, self-anchoring and self-stereotyping research should use an equal balance of positive and negative traits (Otten & Epstude, 2006; van Veelen, Otten, et al., 2016). This research did not do this, but we did control for the social desirability of our positive traits in separate models (M. Cadinu et al., 2013; de la Haye, 2000; Krueger & Clement, 1994) to ensure that the effects were not reducible to a combination of self- and ingroup-enhancement.

The study is also limited in its ability to disambiguate the accuracy of people's group classifications from self-anchoring. Specifically, it may be that some people are in fact relatively similar to their groups and the concordance in similarity-to-self and ingroup classifications is not self-projection as much as accurate classifications. This interpretation can partially be discarded due to the effects identified in Study 1, which involved minimal groups and there should be no ground truth or accurate attributes to classify minimal groups. We further attempted to parse this possibility by estimating and controlling for a variable in separate models that reflects the proportion of instances a given trait was classified as characteristic of the ingroup. This is an imperfect measure as it is implemented on the full sample for which self-to-ingroup generalization is occurring, but it should provide a coarse estimate of average beliefs of how characteristic of the ingroup a given trait is, and controlling for it does not alter inferences. Future research may attempt to further disambiguate these questions of the extent to which people perceive similarity to their ingroup (potentially due to self-stereotyping or self-anchoring)

versus the extent to which people actually are similar to their ingroup (potentially due to acculturation or social customs).

Conclusion

The "*sense of sameness* is the very keel and backbone of our thinking," (James, 1890) and similarity plays a fundamental role in people's everyday reasoning and inferences (Goldstone et al., 1991; Shepard, 1987; Tversky, 1977), including how people reason about themselves and the social groups that they belong to. Using a network-based model (J. Elder, Cheung, et al., 2023) of semantic representation (Griffiths et al., 2007), we introduce model formalism to the long-held tenets of Social Identity Theory (Tajfel, 1978; Turner et al., 1987) regarding the underpinnings of social identification and self-group overlap. We find people use semantic similarity to generalize trait self-beliefs to their ingroup, and that Similarity-to-ingroup primarily drives this effect. People who perceive themselves as more similar to their ingroup, who more strongly identify with their ingroup, or who express greater intergroup bias are more extreme in their tendency to self-project during self-to-ingroup generalization. Moreover, people self-anchor less strongly when they contrast their ingroup identity against a more positively perceived outgroup identity, reflecting that the motivation to differentiate by projecting oneself onto the ingroup may be amplified when contrasting one's ingroup against a more disliked outgroup. Findings further support that people are motivated to belong to their social groups but are not merely chameleons whose self-beliefs are entirely fluctuating to varying social contexts and identity-based cues. Rather, people use the similarities among their various self-beliefs to infer and generalize what ought to be characteristic of their

group as well, which in turn reflects people's various feelings about their social groups. But people do not represent their beliefs about similarity in a vacuum (Nosofsky, 2011), and tendencies to generalize about the self to the ingroup may be amplified under conditions of social tension or conflict (Tajfel, 1974), such as when majority and minority members are pitted against each other. The current work provides important insight into the mechanisms by which "We" becomes "I" and the conditions under which this occurs, while advancing and formalizing prior social psychological theory on the topic with more precise methodology and measurement.

References

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., & Syme, S. L. (1994). Socioeconomic status and health: The challenge of the gradient. *American Psychologist, 49*, 15–24. https://doi.org/10.1037/0003-066X.49.1.15

Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. *Computational Psychiatry (Cambridge, Mass.), 1*, 24–57. https://doi.org/10.1162/CPSY_a_00002

Ahn, W.-Y., Krawitz, A., Kim, W., Busmeyer, J. R., & Brown, J. W. (2011). A Model-Based fMRI Analysis with Hierarchical Bayesian Parameter Estimation. *Journal of Neuroscience, Psychology, and Economics, 4*(2), 95–110. https://doi.org/10.1037/a0020684

Ames, D. R. (2004). Inside the mind reader's tool kit: Projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology, 87*(3), 340–353. https://doi.org/10.1037/0022-3514.87.3.340

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bafumi, J., & Gelman, A. (2007). *Fitting Multilevel Models When Predictors and Group Effects Correlate* (SSRN Scholarly Paper No. 1010095). https://doi.org/10.2139/ssrn.1010095

Balliet, D., Wu, J., & De Dreu, C. K. W. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin, 140*, 1556–1581. https://doi.org/10.1037/a0037737

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 10.1016/j.jml.2012.11.001. https://doi.org/10.1016/j.jml.2012.11.001

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*, 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Bianchi, M., Machunsky, M., Steffens, M. C., & Mummendey, A. (2009). Like me or like us: Is ingroup projection just social projection? *Experimental Psychology, 56*, 198–205. https://doi.org/10.1027/1618-3169.56.3.198

Blake, K. R., & Gangestad, S. (2020). On attenuated interactions, measurement error, and statistical power: Guidelines for social and personality psychologists. *Personality and Social Psychology Bulletin*, *46*, 1702–1711. https://doi.org/10.1177/0146167220913363

Bowman, C. R., Iwashita, T., & Zeithamova, D. (2020). Tracking prototype and exemplar representations in the brain across learning. *ELife*, *9*, e59360. https://doi.org/10.7554/eLife.59360

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, *17*(5), 475–482.

Brewer, M. B. (1993). Social identity, distinctiveness, and in-group homogeneity. *Social Cognition*, *11*, 150–164. https://doi.org/10.1521/soco.1993.11.1.150

Brewer, M. B. (1999). The Psychology of Prejudice: Ingroup Love and Outgroup Hate? *Journal of Social Issues*, *55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Brown, R. (2000). Social identity theory: Past achievements, current problems and future challenges. *European Journal of Social Psychology*, *30*(6), 745–778. https://doi.org/10.1002/1099-0992(200011/12)30:6<745::AID-EJSP24>3.0.CO;2-O

Brown, R. J. (1984). The role of similarity in intergroup relations. In H. Tajfel (Ed.), *The Social Dimension: European Developments in Social Psychology* (Vol. 2, pp. 603–623). Cambridge University Press. https://doi.org/10.1017/CBO9780511759154.012

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116.

Cadinu, M., Carnaghi, A., & Guizzo, F. (2020). Group meaningfulness and the causal direction of influence between the ingroup and the self or another individual: Evidence from the Induction-Deduction Paradigm. *PLoS ONE*, *15*(3), e0229321. https://doi.org/10.1371/journal.pone.0229321

Cadinu, M., & De Amicis, L. (1999). The relationship between the self and the ingroup: When having a common conception helps. *Swiss Journal of Psychology / Schweizerische Zeitschrift Für Psychologie / Revue Suisse de Psychologie*, *58*, 226–232. https://doi.org/10.1024/1421-0185.58.4.226

Cadinu, M., Latrofa, M., & Carnaghi, A. (2013). Comparing Self-stereotyping with In-group-stereotyping and Out-group-stereotyping in Unequal-status Groups: The

Case of Gender. *Self and Identity*, *12*(6), 582–596. https://doi.org/10.1080/15298868.2012.712753

Cadinu, M. R., & Rothbart, M. (1996). Self-anchoring and differentiation processes in the minimal group setting. *Journal of Personality and Social Psychology*, *70*(4), 661.

Cadinu, M., & Rothbart, M. (1996). Self-anchoring and differentiation processes in the minimal group setting. *Journal of Personality and Social Psychology*, *70*, 661–677. https://doi.org/10.1037/0022-3514.70.4.661

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science*, *3*(1), 14–25. https://doi.org/10.1002/bs.3830030103

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology*, *70*(1), 141–156. https://doi.org/10.1037/0022-3514.70.1.141

Clement, R. W., & Krueger, J. (2000). The primacy of self-referent information in perceptions of social consensus. *British Journal of Social Psychology*, *39*, 279–299. https://doi.org/10.1348/014466600164471

Clement, R. W., & Krueger, J. (2002). Social Categorization Moderates Social Projection. *Journal of Experimental Social Psychology*, *38*(3), 219–231. https://doi.org/10.1006/jesp.2001.1503

Coats, S., Smith, E. R., Claypool, H. M., & Banner, M. J. (2000). Overlapping mental representations of self and in-group: Reaction time evidence and its relationship with explicit measures of group indentification. *Journal of Experimental Social Psychology*, *36*, 304–315. https://doi.org/10.1006/jesp.1999.1416

Davis, T., & Goldwater, M. (2021). Using model-based neuroimaging to adjudicate structured and continuous representational accounts in same-different categorization and beyond. *Current Opinion in Behavioral Sciences*, *37*, 103–108. https://doi.org/10.1016/j.cobeha.2020.11.011

Davis, T., & Love, B. C. (2010). Memory for Category Information Is Idealized Through Contrast With Competing Options. *Psychological Science*, *21*(2), 234–242. https://doi.org/10.1177/0956797609357712

Davis, T., Love, B. C., & Preston, A. R. (2012a). Striatal and Hippocampal Entropy and Recognition Signals in Category Learning: Simultaneous Processes Revealed by Model-Based fMRI. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(4), 821–839. https://doi.org/10.1037/a0027865

Davis, T., Love, B. C., & Preston, A. R. (2012b). Learning the Exception to the Rule: Model-Based fMRI Reveals Specialized Representations for Surprising Category Members. *Cerebral Cortex*, *22*(2), 260–273. https://doi.org/10.1093/cercor/bhr036

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. In *Decision making, affect, and learning: Attention and performance XXIII* (1st ed., Vol. 23). Oxford University Press Oxford.

de la Haye, A.-M. (2000). A methodological note about the measurement of the false-consensus effect. *European Journal of Social Psychology*, *30*(4), 569–581. https://doi.org/10.1002/1099-0992(200007/08)30:4<569::AID-EJSP8>3.0.CO;2-V

DiDonato, T. E., Ullrich, J., & Krueger, J. I. (2011). Social perception as induction and inference: An integrative model of intergroup differentiation, ingroup favoritism, and differential accuracy. *Journal of Personality and Social Psychology*, *100*, 66–83. https://doi.org/10.1037/a0021051

Don, H. J., Otto, A. R., Cornwall, A. C., Davis, T., & Worthy, D. A. (2019). Learning reward frequency over reward probability: A tale of two learning rules. *Cognition*, *193*, 104042. https://doi.org/10.1016/j.cognition.2019.104042

Don, H. J., & Worthy, D. A. (2022). Frequency effects in action versus value learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*, 1311–1327. https://doi.org/10.1037/xlm0000896

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

Eager, C., & Roy, J. (2017). *Mixed Effects Models are Sometimes Terrible* (arXiv:1701.04858). arXiv. https://doi.org/10.48550/arXiv.1701.04858

Elder, G. H. (1998). The Life Course as Developmental Theory. *Child Development*, *69*(1), 1–12. https://doi.org/10.2307/1132065

Elder, J., Cheung, B., Davis, T., & Hughes, B. (2023). Mapping the self: A network approach for understanding psychological and neural representations of self-concept structure. *Journal of Personality and Social Psychology*, *124*(2), 237–263. https://doi.org/10.1037/pspa0000315

Elder, J., Davis, T. H., & Hughes, B. L. (2023a). A fluid self-concept: How the brain maintains coherence and positivity across an interconnected self-concept while

incorporating social feedback. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1951-22.2023

Elder, J., Davis, T., & Hughes, B. (2022a). *Self-Concept Certainty and Stability are Associated with Semantic Relations Between Self-Beliefs*. PsyArXiv. https://doi.org/10.31234/osf.io/8y2vr

Elder, J., Davis, T., & Hughes, B. (2022b). *Self-derogating to align with group norms: Similarity-based learning drives assimilation of the group to the self*. PsyArXiv. https://doi.org/10.31234/osf.io/38bf7

Elder, J., Davis, T., & Hughes, B. (2023b). A fluid self-concept: How the brain maintains coherence and positivity across an interconnected self-concept while incorporating social feedback. *The Journal of Neuroscience*.

Elder, J., Davis, T., & Hughes, B. L. (2022c). Learning About the Self: Motives for Coherence and Positivity Constrain Learning From Self-Relevant Social Feedback. *Psychological Science*, *33*(4), 629–647. https://doi.org/10.1177/09567976211045934

Ellemers, N., Spears, R., & Doosje, B. (2001). *SELF AND SOCIAL IDENTITY*. 27.

Estes, W. K. (1976a). The cognitive side of probability learning. *Psychological Review*, *83*, 37–64. https://doi.org/10.1037/0033-295X.83.1.37

Estes, W. K. (1976b). Some Functions of Memory in Probability Learning and Choice Behavior11The research reported here was supported primarily by USPHS grant MH16100 from the National Institute of Mental Health; the theoretical analyses and the preparation of this chapter were supported also by grant MH23878. I wish to acknowledge the contribution of Edith Skaar to the data collection and processing. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 10, pp. 1–45). Academic Press. https://doi.org/10.1016/S0079-7421(08)60463-6

FeldmanHall, O., & Shenhav, A. (2019). Resolving uncertainty in a social world. *Nature Human Behaviour*, *3*(5), Article 5. https://doi.org/10.1038/s41562-019-0590-x

Fielding, K. S., & Hogg, M. A. (1997). Social identity, self-categorization, and leadership: A field study of small interactive groups. *Group Dynamics: Theory, Research, and Practice*, *1*(1), 39–51. https://doi.org/10.1037/1089-2699.1.1.39

Fiske, S. T., Dupree, C. H., Nicolas, G., & Swencionis, J. K. (2016). Status, power, and intergroup relations: The personal is the societal. *Current Opinion in Psychology*, *11*, 44–48. https://doi.org/10.1016/j.copsyc.2016.05.012

Gaertner, S. L., Dovidio, J. F., Guerra, R., Hehman, E., & Saguy, T. (2016). A common ingroup identity: Categorization, identity, and intergroup relations. In *Handbook of prejudice, stereotyping, and discrimination, 2nd ed* (pp. 433–454). Psychology Press. https://doi.org/10.4324/9781841697772

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology*, *23*(2), 222–262. https://doi.org/10.1016/0010-0285(91)90010-L

Gramzow, R. H., & Gaertner, L. (2005). Self-Esteem and Favoritism Toward Novel In-Groups: The Self as an Evaluative Base. *Journal of Personality and Social Psychology*, *88*(5), 801–815. https://doi.org/10.1037/0022-3514.88.5.801

Gramzow, R. H., Gaertner, L., & Sedikides, C. (2001). Memory for in-group and out-group information in a minimal group context: The self as an informational base. *Journal of Personality and Social Psychology*, *80*(2), 188–205. https://doi.org/10.1037/0022-3514.80.2.188

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*, 211–244. https://doi.org/10.1037/0033-295X.114.2.211

Guest, O., & Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science*, *16*(4), 789–802. https://doi.org/10.1177/1745691620970585

Guillaume, J. H. A., Jakeman, J. D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., Hill, M. C., Jakeman, A. J., Keesman, K. J., Razavi, S., & Stigter, J. D. (2019). Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose. *Environmental Modelling & Software*, *119*, 418–432. https://doi.org/10.1016/j.envsoft.2019.07.007

Haslam, S. A., Oakes, P. J., Turner, J. C., & McGarty, C. (1995). Social categorization and group homogeneity: Changes in the perceived applicability of stereotype content as a function of comparative context and trait favourableness. *British Journal of Social Psychology*, *34*(2), 139–160. https://doi.org/10.1111/j.2044-8309.1995.tb01054.x

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *20*(2), 411–422. https://doi.org/10.1037//0278-7393.20.2.411

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575–604. https://doi.org/10.1146/annurev.psych.53.100901.135109

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychological Review*, *119*(2), 304–320. https://doi.org/10.1037/a0026767

Hogg, M. A. (2007). Uncertainty-identity theory. In *Advances in experimental social psychology, Vol 39* (pp. 69–126). Elsevier Academic Press. https://doi.org/10.1016/S0065-2601(06)39002-8

Hogg, M. A. (2014). From Uncertainty to Extremism: Social Categorization and Identity Processes. *Current Directions in Psychological Science*, *23*(5), 338–342. https://doi.org/10.1177/0963721414540168

Hogg, M. A., & Abrams, D. (1988). *Social identifications: A social psychology of intergroup relations and group processes* (pp. xv, 268). Taylor & Frances/Routledge.

Hogg, M. A., Abrams, D., Otten, S., & Hinkle, S. (2004). The Social Identity Perspective: Intergroup Relations, Self-Conception, and Small Groups. *Small Group Research*, *35*(3), 246–276. https://doi.org/10.1177/1046496404263424

Hogg, M. A., & Hains, S. C. (1996). Intergroup relations and group solidarity: Effects of group identification and social beliefs on depersonalized attraction. *Journal of Personality and Social Psychology*, *70*(2), 295–309. https://doi.org/10.1037/0022-3514.70.2.295

Hogg, M. A., Hardie, E. A., & Reynolds, K. J. (1995). Prototypical similarity, self-categorization, and depersonalized attraction: A perspective on group cohesiveness. *European Journal of Social Psychology*, *25*(2), 159–177. https://doi.org/10.1002/ejsp.2420250204

Hong, Y., & Ratner, K. G. (2021). Minimal but not meaningless: Seemingly arbitrary category labels can imply more than group membership. *Journal of Personality and Social Psychology*, *120*(3), 576–600. https://doi.org/10.1037/pspa0000255

Hornsey, M. J., & Hogg, M. A. (1999). Subgroup differentiation as a response to an overly-inclusive group: A test of optimal distinctiveness theory. *European*

Journal of Social Psychology, *29*, 543–550. https://doi.org/10.1002/(SICI)1099-0992(199906)29:4<543::AID-EJSP945>3.0.CO;2-A

Hornsey, M. J., & Hogg, M. A. (2000). Assimilation and diversity: An integrative model of subgroup relations. *Personality and Social Psychology Review*, *4*, 143–156. https://doi.org/10.1207/S15327957PSPR0402_03

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. *PLOS Computational Biology*, *7*(4), e1002028. https://doi.org/10.1371/journal.pcbi.1002028

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science*, *22*(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

James, W. (1890). *The principles of psychology, Vol I.* (pp. xii, 697). Henry Holt and Co. https://doi.org/10.1037/10538-000

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, *33*(1), 1–27. https://doi.org/10.1111/j.2044-8309.1994.tb01008.x

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Krueger, J. (2007). From social projection to social behaviour. *European Review of Social Psychology*, *18*(1), 1–35. https://doi.org/10.1080/10463280701284645

Krueger, J., & Clement, R. W. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology*, *67*, 596–610. https://doi.org/10.1037/0022-3514.67.4.596

Krueger, J., & Clement, R. W. (1996). Inferring category characteristics from sample characteristics: Inductive reasoning and social projection. *Journal of Experimental Psychology: General*, *125*(1), 52.

Kubinec, R. (2022). Ordered Beta Regression: A Parsimonious, Well-Fitting Model for Continuous Data with Lower and Upper Bounds. *Political Analysis*, 1–18. https://doi.org/10.1017/pan.2022.20

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Latrofa, M., Vaes, J., & Cadinu, M. (2012). Self-Stereotyping: The Central Role of an Ingroup Threatening Identity. *The Journal of Social Psychology*, *152*(1), 92–111. https://doi.org/10.1080/00224545.2011.565382

Latrofa, M., Vaes, J., Pastore, M., & Cadinu, M. (2009). "United we stand, divided we fall'! The protective function of self-stereotyping for stigmatised members' psychological well-being. *Applied Psychology: An International Review*, *58*(1), 84–104. https://doi.org/10.1111/j.1464-0597.2008.00383.x

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

Leary, M. R., Kelly, K. M., Cottrell, C. A., & Schreindorfer, L. S. (2013). Construct Validity of the Need to Belong Scale: Mapping the Nomological Network. *Journal of Personality Assessment*, *95*(6), 610–624. https://doi.org/10.1080/00223891.2013.819511

Lelkes, Y., & Westwood, S. J. (2017). The Limits of Partisan Prejudice. *The Journal of Politics*, *79*(2), 485–501. https://doi.org/10.1086/688223

Leonardelli, G. J., Pickett, C. L., & Brewer, M. B. (2010). Optimal distinctiveness theory: A framework for social identity, social cognition, and intergroup relations. In *Advances in experimental social psychology* (Vol. 43, pp. 63–113). Elsevier.

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Luhtanen, R., & Crocker, J. (1992). A Collective Self-Esteem Scale: Self-Evaluation of One's Social Identity. *Personality and Social Psychology Bulletin*, *18*(3), 302–318. https://doi.org/10.1177/0146167292183006

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*(1), 49–70. https://doi.org/10.3758/BF03211715

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of Effect Existence and Significance in the Bayesian Framework. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02767

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, *38*, 299–337. https://doi.org/10.1146/annurev.ps.38.020187.001503

Martin, R. C. (2007). Semantic short-term memory, language processing, and inhibition. In *Automaticity and control in language processing* (pp. 161–191). Psychology Press.

Mullen, B. (1991). Group composition, salience, and cognitive representations: The phenomenology of being in a group. *Journal of Experimental Social Psychology*, *27*, 297–323. https://doi.org/10.1016/0022-1031(91)90028-5

Murayama, K., Usami, S., & Sakaki, M. (2022). Summary-statistics-based power analysis: A new and practical method to determine sample size for mixed-effects modeling. *Psychological Methods*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/met0000330

Nelson, L. J., & Miller, D. T. (1995). The distinctiveness effect in social categorization: You are what makes you unusual. *Psychological Science*, *6*, 246–249. https://doi.org/10.1111/j.1467-9280.1995.tb00600.x

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas—Part II. *Language and Linguistics Compass*, *10*(11), 591–613. https://doi.org/10.1111/lnc3.12207

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 104–114. https://doi.org/10.1037/0278-7393.10.1.104

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 54–65. https://doi.org/10.1037/0278-7393.14.1.54

Nosofsky, R. M. (1991). Typicality in logically defined categories: Exemplar-similarity versus rule instantiation. *Memory & Cognition*, *19*(2), 131–150. https://doi.org/10.3758/BF03197110

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In A. J. Wills & E. M. Pothos (Eds.), *Formal Approaches in Categorization* (pp. 18–39). Cambridge University Press. https://doi.org/10.1017/CBO9780511921322.002

Obst, P., White, K., Mavor, K., & Baker, R. (2011). Social Identification Dimensions as Mediators of the Effect of Prototypicality on Intergroup Behaviours. *Psychology, 2*. https://doi.org/10.4236/psych.2011.25066

Otten, S. (2003). "Me and us" or "us and them"? The self as a heuristic for defining minimal ingroups. *European Review of Social Psychology*, *13*(1), 1–33. https://doi.org/10.1080/10463280240000028

Otten, S. (2004). Self-anchoring as predictor of in-group favoritism: Is it applicable to real group contexts? *Current Psychology of Cognition*, *22*(4/5), 427.

Otten, S., & Epstude, K. (2006). Overlapping mental representations of self, ingroup, and outgroup: Unraveling self-stereotyping and self-anchoring. *Personality & Social Psychology Bulletin*, *32*(7), 957–969. https://doi.org/10.1177/0146167206287254

Otten, S., & Wentura, D. (2001). Self-Anchoring and In-Group Favoritism: An Individual Profiles Analysis. *Journal of Experimental Social Psychology*, *37*(6), 525–532. https://doi.org/10.1006/jesp.2001.1479

Postmes, T., Haslam, S. A., & Jans, L. (2013). A single-item measure of social identification: Reliability, validity, and utility. *British Journal of Social Psychology*, *52*(4), 597–617. https://doi.org/10.1111/bjso.12006

Rabinovich, A., & Morton, T. A. (2011). Subgroup identities as a key to cooperation within large social groups. *British Journal of Social Psychology*, *50*(1), 36–51. https://doi.org/10.1348/014466610X486356

Riketta, M., & Sacramento, C. A. (2008). `They Cooperate With Us, So They Are Like Me': Perceived Intergroup Relationship Moderates Projection from Self to Outgroups. *Group Processes & Intergroup Relations*, *11*(1), 115–131. https://doi.org/10.1177/1368430207084849

Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible Hands and Fine Calipers: A Call to Use Formal Theory as a Toolkit for Theory Construction. *Perspectives on Psychological Science*, *16*(4), 725–743. https://doi.org/10.1177/1745691620974697

Rosch, E., & Lloyd, B. B. (1978). *Principles of categorization*.

Rosenberg, M. (1965). Rosenberg self-esteem scale (RSE). *Acceptance and Commitment Therapy. Measures Package*, *61*(52), 18.

Roth, J., Steffens, M. C., & Vignoles, V. L. (2018). Group Membership, Group Change, and Intergroup Attitudes: A Recategorization Model Based on Cognitive Consistency Principles. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00479

Schneider, M. J., Rubin-McGregor, J., Elder, J., Hughes, B., & Tamir, D. (2022). *Simulation requires activation of self-knowledge to change self-concept.* PsyArXiv. https://doi.org/10.31234/osf.io/92mru

Schubert, T. W., & Otten, S. (2002). Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization. *Self and Identity*, *1*, 353–376. https://doi.org/10.1080/152988602760328012

Sell, J., & Love, T. P. (2009). Common fate, crisis, and cooperation in social dilemmas. In S. R. Thye & E. J. Lawler (Eds.), *Altruism and Prosocial Behavior in Groups* (Vol. 26, pp. 53–79). Emerald Group Publishing Limited. https://doi.org/10.1108/S0882-6145(2009)0000026006

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shepard, R. N. (1987). Toward a Universal Law of Generalization for Psychological Science. *Science*, *237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243

Sherman, D. K., & Kim, H. S. (2005). Is There an "I" in "Team"? The Role of the Self in Group-Serving Judgments. *Journal of Personality and Social Psychology*, *88*, 108–120. https://doi.org/10.1037/0022-3514.88.1.108

Simon, B., & Brown, R. (1987). Perceived intragroup homogeneity in minority-majority contexts. *Journal of Personality and Social Psychology*, *53*, 703–711. https://doi.org/10.1037/0022-3514.53.4.703

Smith, E. R., & Henry, S. (1996). An in-group becomes part of the self: Response time evidence. *Personality and Social Psychology Bulletin*, *22*, 635–642. https://doi.org/10.1177/0146167296226008

Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology*, *12*(3), 175–200. https://doi.org/10.20982/tqmp.12.3.p175

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology, 15*, 201–293. https://doi.org/10.2307/1412107

Spears, R., Doosje, B., & Ellemers, N. (1997). Self-Stereotyping in the Face of Threats to Group Status and Distinctiveness: The Role of Group Identification. *Personality and Social Psychology Bulletin*, *23*(5), 538–553. https://doi.org/10.1177/0146167297235009

Spencer-Rodgers, J., Srivastava, S., Boucher, H. C., English, T., Paletz, S. B., & Peng, K. (2015). The dialectical self scale. *Unpublished Manuscript*. https://doi.org/10.13140/RG.2.1.3695.5928

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information*, *13*(2), 65–93. https://doi.org/10.1177/053901847401300204

Tajfel, H. (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations.* Academic Press.

Tajfel, H., & Billig, M. (1974). Familiarity and categorization in intergroup behavior. *Journal of Experimental Social Psychology*, *10*, 159–170. https://doi.org/10.1016/0022-1031(74)90064-X

Tajfel, H., Sheikh, A. A., & Gardner, R. C. (1964). Content of stereotypes and the inference of similarity between members of stereotyped groups. *Acta Psychologica*.

Tropp, L. R., & Wright, S. C. (2001). Ingroup identification as the inclusion of ingroup in the self. *Personality and Social Psychology Bulletin*, *27*, 585–600. https://doi.org/10.1177/0146167201275007

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory* (pp. x, 239). Basil Blackwell.

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

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. https://doi.org/10.1037/0033-295X.84.4.327

Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, *5*, 244–252. https://doi.org/10.1016/S1364-6613(00)01651-X

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 219–230. https://doi.org/10.1037/0278-7393.33.1.219

Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by Repetition: Explanations and Implications. *Current Directions in Psychological Science*, *28*(3), 247–253. https://doi.org/10.1177/0963721419827854

Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, *160*, 110–126. https://doi.org/10.1016/j.cognition.2016.12.016

van Veelen, R., Eisenbeiss, K. K., & Otten, S. (2016). Newcomers to Social Categories: Longitudinal Predictors and Consequences of Ingroup Identification. *Personality and Social Psychology Bulletin*, *42*(6), 811–825. https://doi.org/10.1177/0146167216643937

van Veelen, R., Otten, S., Cadinu, M., & Hansen, N. (2016). An Integrative Model of Social Identification: Self-Stereotyping and Self-Anchoring as Two Cognitive Pathways. *Personality and Social Psychology Review*, *20*(1), 3–26. https://doi.org/10.1177/1088868315576642

van Veelen, R., Otten, S., & Hansen, N. (2011). Linking self and ingroup: Self-anchoring as distinctive cognitive route to social identification. *European Journal of Social Psychology*, *41*(5), 628–637. https://doi.org/10.1002/ejsp.792

van Veelen, R., Otten, S., & Hansen, N. (2013a). Social identification when an in-group identity is unclear: The role of self-anchoring and self-stereotyping. *British Journal of Social Psychology*, *52*(3), 543–562. https://doi.org/10.1111/j.2044-8309.2012.02110.x

van Veelen, R., Otten, S., & Hansen, N. (2013b). A personal touch to diversity: Self-anchoring increases minority members' identification in a diverse group. *Group Processes & Intergroup Relations*, *16*(6), 671–683. https://doi.org/10.1177/1368430212473167

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved Rˆ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, *16*(2), 667–718. https://doi.org/10.1214/20-BA1221

Vogel, T., Carr, E. W., Davis, T., & Winkielman, P. (2018). Category structure determines the relative attractiveness of global versus local averages. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 250–267. https://doi.org/10.1037/xlm0000446

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1. https://doi.org/10.1017/S0140525X20001685

Zeithamova, D., Maddox, W. T., & Schnyer, D. M. (2008). Dissociable Prototype Learning Systems: Evidence from Brain Imaging and Behavior. *Journal of Neuroscience*, *28*(49), 13194–13201. https://doi.org/10.1523/JNEUROSCI.2915-08.2008