Mere Membership

Yarrow Dunham*
Department of Psychology
Yale University

* Correspondence: yarrow.dunham@yale.edu (Yarrow Dunham)

Keywords: intergroup cognition; prejudice; minimal groups

Accepted manuscript, Trends in Cognitive Sciences
Abstract

Human social groups are central to social organization and pervasively impact interpersonal interactions. While immensely varied, all social groups can also be considered specific instantiations of a common and abstract ingroup-outgroup structure. How much of the power of human social groups stems from learned variation versus abstract commonality? I review evidence demonstrating that from early in development a wide range of intergroup phenomena, most prominently many ingroup biases, follow solely from simple membership in an abstract social collective. Such effects cannot be attributed to rich social learning and so: a) constrain theories seeking to explain or intervene on ingroup bias, and b) provide reason to think that our species is powerfully predisposed towards ingroup favoritism from early in development.
Centrality of the group and the prevalence of ingroup bias

No account of human psychology would be complete without recognizing that other people are as central to our natural ecology as is the physical environment. In particular, social groups—collections of people united by commonality—form the backdrop against which human culture emerges in all its diversity. While many species recognize a small set of social groups such as sex, kin, and band [1,2], human social groupings are enigmatic in their variety. We not only accept but embrace, act upon, and structure social arrangements around a range of groups that on the face of it have almost nothing in common. Consider, for example, gender, race, ethnicity, religion, occupation, nationality, language. The variation encompassed by these categories becomes clear when we attend to the properties that define them: socio-biological facts, ancestral origin, customs, ideological commitments, skills and roles, contingent aspects of life history, regions of upbringing and accidents of birth. These differences loom large because what follows from (e.g.) speaking the same language is quite different from what follows from (e.g.) sharing a political ideology. At first glance, this variability would seem to make mastering this conceptual space a challenging learning problem. But if we zoom out from this variability it gives way to a common form of social geography: every group marks a category to which one does or does not belong, a division into “us” and “them.” And here we arrive at a disturbing fact: Each of these dimensions of personhood (and many more besides) tends to produce ingroup bias, a constellation of psychological tendencies that favor the groups to which we belong. Given its intimate relationship to discrimination and conflict, research on ingroup bias has been a mainstay of the social sciences.

What explains the pervasive presence of ingroup bias? Proposed explanations are numerous, spanning multiple levels of analysis. Some of the most prominent: genuine conflict over scarce resources [3,4]; an evolved coalitional psychology [5,6]; epistemic uncertainty which can be assuaged by the security of the tribe [7-9]; rich social or personality motivations to bolster and positively differentiate the self [10-13]; the transmission and replication of biased portrayals of groups [14-16]; a suite of moral norms urging ingroup loyalty [15]. This proliferation of accounts is at least in part due to the fact that what must be explained—the range of phenomena that can be called “ingroup bias”—is itself so vast. In addition to merely preferring the ingroup over the outgroup, biases appear in many domains including empathy [17], (de)humanization
[18], mind perception [19], moral attribution [20], cooperation [21], and generosity [22]. Some of these effects are observed with respect to many social groups (for example merely liking the ingroup more), while others appear to target some groups in particular (for example finding an outgroup particularly threatening [23]). Different theoretical perspectives, brought to bear to explain different phenomena, have made the search for a unified account elusive.

The goal of this review is to impose some clarity on a crowded field by synthesizing research demonstrating that many and perhaps even most forms of ingroup bias emerge as the spontaneous consequence of self-categorization into social groups. Because they emerge early in development and have a pervasive effect on social evaluation and learning, these phenomena strongly suggest that intergroup cognition in humans is from its inception powerfully channeled towards ingroup favoritism. And because they emerge even when groups are devoid of richer meaning, these findings impose under-appreciated constraints on the range of plausible accounts of ingroup bias, suggesting that a number of putative causal ingredients are in fact not necessary for the emergence of ingroup bias or many of its consequences. Taken seriously, these considerations also support a programmatic roadmap for future research with the goal of carefully delineating what does and does not follow from the simple act of self-categorization.

Mere membership and its consequences

There is intuitive appeal in marshalling rich and varied theoretical resources to explain rich and varied forms of human sociality. But what is most striking about ingroup bias is not its occurrence in the midst of that richness. Rather, what is most striking is its occurrence in highly generalized and impersonal contexts in which many of the candidate causes just referenced are minimized or absent. Indeed, ingroup bias appears even absent conflict, competition, or indeed any information whatsoever about the specific characteristics of the groups involved. The red-banner example is the minimal groups paradigm [24,25] (MGP; Box 1). While it admits of many variants, in its simplest and clearest instantiation isolated individuals are randomly assigned to previously unfamiliar social groups based on arbitrary cues such as a group-denoting color or novel label. This manipulation, a social coin flip, has profound consequences, producing a wide range of biases in favor of the ingroup [25]. While the MGP is widely known, its consequences are broader than generally recognized, making it paradoxically both familiar and
underappreciated. Indeed, a central goal of this review is to detail some fundamental challenges these effects pose to common accounts of the origins of ingroup bias [25-27].

Let us call the primary outcome of the MGP “mere membership”, the situation in which you know you belong to a group (having been so assigned), but as yet know nothing else about the group. The consequences of this superficially meaningless state of affairs are impressively broad: mere membership spontaneously elicits literally dozens of forms of ingroup-favoritism across evaluative, coalitional, and learning domains, many of which emerge early in childhood [27-29]. What follows from mere membership is therefore not an effect, but rather a rich constellation or syndrome of ingroup-favoring cognitions, of which Table 1 (Key Table) provides a detailed summary.

Because the effects of mere membership cannot be attributed to anything specifically learned about one group or another, and because they emerge well before children develop detailed knowledge about most of the groups that structure their world, the most interesting possibility is that these effects illuminate—or even constitute—the abstract structure of our thinking about groups, a template over which subsequent learning occurs. Children emerge into highly variable cultural worlds in which quite different social distinctions hold sway. The challenge of mastering and coming to properly value multiple social distinctions could be abetted by a set of basic assumptions about group function: when a new intergroup distinction is encountered, it would only need to be mapped onto this pre-existing template to immediately engage the myriad consequences described in Table 1. While still speculative, support for this possibility comes from multiple lines of recent work. First is adult work demonstrating a close correspondence between how real and minimal groups function. For example, most of the effects summarized in Table 1 emerge in similar form with respect to at least some culturally meaningful groups; that is, membership in real and minimal groups often has quite similar cognitive and affective consequences. Higher-order commonalities also exist. Cognitively, ingroup bias in both real and minimal groups can be described as a form of cognitive consistency or “balance” between group identification and self-esteem, such that manipulations of any one “leg” of this cognitive-affective triangle affect the other two [30,31]. And neuroscientifically, recent work has identified distinct brain signatures of intergroup categorization [26,32]. In one such study, individuals were placed into minimal groups and a classifier was trained on patterns of brain activation from the
incidental observation of minimal ingroup and outgroup targets; that classifier was than tested on brain activation in response to political ingroup and outgroup targets, and was able to identify group membership along this orthogonal dimension with high accuracy [26]. Thus, at multiple levels of analysis, from individual phenomena through more general aspects of cognitive and neural structure, minimal groups seem capable of explaining many aspects of intergroup functioning.

Recent developmental work further enriches this picture, in particular studies with infants and young children that demonstrate sophisticated intuitions about how groups function. For example, infants favor agents that share their preferences, even newly elicited preferences [33], and slightly older children favor minimally-defined ingroup members and expect those ingroup members to be similar to them [27]. Further, infants observing third-party interactions expect members of minimally defined groups to act in similar ways [34], and to enact basic principles of coalitional interaction, including aiding fellow group members, especially under conditions of resource scarcity [35,36].

It is thus tempting to consider this abstract knowledge about and responsiveness to groups as an evolved and early-emerging response to our deep embeddedness in both enduring and rapidly shifting coalitions [5,37-39]. Additional tests of this possibility are readily imaginable, for example linking the minimal group effect to bias towards real-world groups, or perhaps more powerfully, towards so-called “generalized prejudice” estimated from an individual’s attitudes towards multiple groups [13,40]. But even if this admittedly rich interpretation fails to convince, the weight of the empirical work demonstrating the consequences of membership in minimal groups still must be reckoned with because it raises a set of challenges to the ways in which intergroup phenomena are frequently conceptualized. It is to these challenges that we now turn.

**Two challenges: Underlying mechanisms and learned specificity**

Two related dimensions of complexity are central to the science of intergroup cognition. First is the range of causal factors marshalled to explain the emergence of ingroup bias. Second is the range of intergroup phenomena itself, many appearing to implicate some specific intergroup targets, and therefore presupposing specific learning histories (for example, information
characterizing racial or religious categories in North America). However, the breadth of what follows from mere membership presents a clarifying challenge to each of these dimensions.

Beginning with putative causal factors, the MGP demonstrates that mere membership is sufficient to elicit many forms of ingroup bias. This sufficiency threatens some other claims of causal necessity: if a potential cause of a form of ingroup bias that follows from mere membership is not itself present in the MGP, that cause cannot claim to be necessary. Examples will clarify. Perceptions of competition and conflict have been appealed to as causes of attitudinal and resource-based ingroup bias [3]. But attitudinal and resource-based bias are induced by the MGP, in which there is no conflict or competition. Thus, past work taken as evidence for the causal necessity of those factors may be better understood as evidence that they increase rather than create ingroup bias. Similarly, exposure to positive versus negative representations of groups is frequently appealed to as a cause of ingroup bias [15,16]. But such information is also absent in the minimal groups setting, suggesting that while it surely can affect attitudes, it too cannot make a claim to initial causal necessity.

Other proposed causal processes are, or might be, present in the minimal groups setting. These candidates have a special status as potential explanations for the power of mere membership itself. For example, social identification, motivated uncertainty reduction, or personality processes could operate in the minimal groups setting and so could be (indeed have been proposed as) explanations for minimal group effects themselves [41-43]. The upshot is that if we want to understand what psychological mechanisms lead to ingroup bias, we should begin with the question of what psychological mechanisms are plausibly present in the minimal groups setting. These are the factors we should consider potential causes, while those that are not should be treated separately, as factors that may play a moderating role. Thus, the robust phenomena elicited by the MGP imposes constraints on the factors that can make a plausible claim to causal necessity.

The broad impact of mere membership supports a second, related, challenge. Many intergroup phenomena are assumed to uniquely target some groups but not others. For example, many effects are thought to emerge from the specific dynamics of race in North America. But if a phenomenon emerges from mere membership in an arbitrary social group (a highly general setting) then a claim of specificity is unwarranted. Put differently, every specific real-world group
is at once a group in general (involving, for example, membership) and a group in situ (with a unique socio-cultural history). We want to know which effects arising from membership in a given social category are a consequence of the group merely being a group versus a consequence of something more specific, say the group being race in North America. Unfortunately, telling which is which is not so easy. Box 2 details several cases in which effects widely interpreted as resting on specific culturally embedded learning histories [44,45] turn out to extend in similar fashion to minimal groups [46,47]. These kinds of examples demonstrate that our intuitions are not particularly reliable guides to whether a given effect is specific or general: many things that were plausibly specific turn out to be surprisingly general. It follows that the burden of proof should generally fall on those making claims or assumptions of specificity, because in many cases such assumptions do not hold. Luckily, this is easily done, at least in principle: merely show that the effect of interest does not appear in a sufficiently similar minimal group experiment.

Thus, while it has not always been thought of in this way, the MGP provides both the demonstration of generality as well as the methodological tool which allows us to test the extent of that generality. To determine whether a particular candidate cause is necessary, see if it is plausibly present in the MGP, or if it can be linked to variation in the strength of minimal group preferences. To show that a given effect is specific, show that it does not replicate in the MGP. In this way our explanatory efforts can be partitioned into two parts: elucidating the mechanisms underlying the effects of mere membership, and identifying the effects which are beyond its scope or which moderate its effects. Doing this careful work is all the more critical because dozens of new minimal group experiments are conducted each year, with the cumulative effect of gradually broadening (if also revealing limits of) the known consequences of mere membership.

It follows that a powerful way to enhance our understanding of intergroup bias is to more systematically test the consequences of mere membership, as well as its relationship to biases towards real-world groups. This investigation is particularly timely in light of recent calls for reproducibility and the publication of theoretically relevant null results [46,47]. Indeed, failures to find a particular effect in the MGP are as informative as successes, and so should readily attract attention from researchers in the area. In a similar vein, the present perspective can be taken as encouragement for researchers to open their file drawers to call attention to informative past results, whether in the form of positive or negative findings.
Of course, the argument presented here should not be taken as a dismissal of important work showing how other factors influence ingroup bias, nor of efforts to capture as much of the variance in intergroup phenomena as possible via more complex theoretical edifices. But the effort to clearly delineate necessary and sufficient factors is nonetheless critical as a means of imposing order on a crowded field, separating the factors that directly produce ingroup bias from the factors that exaggerate or attenuate it. And even beyond its importance for theory building, successful causal identification will also benefit intervention efforts. By way of example, invoking rich forms of social learning to explain the effect of anger on racial categorization described in Box 2 should strike us as superfluous or even misleading, because it obscures a clear view of the causal mechanisms involved. Simply put, changing patterns of cultural input is unlikely to effectively reduce a bias that emerges just as powerfully in the wholesale absence of that input.

**Causes of membership effects**

As always, it is important to distinguish levels at which an explanation for a phenomenon is pitched. When it is asserted that at least some important forms of ingroup bias reflect an evolved system of coalitional reasoning, this ultimate explanation as yet says little about the proximal mechanisms implementing it. Conversely, when a theorist suggests that ingroup bias emerges as self-related positivity is extended to the group, this proximal explanation could occur for many reasons, e.g. as a mere byproduct of representational structure or because the link between self and group became psychologically privileged over evolutionary time. While our focus here will be on proposed psychological mechanisms, it will behoove us to keep an eye on whether the mechanism is conceived of as functional (a learned or evolved strategy to solve a specific set of problems) or merely a byproduct of other psychological processes (such as those relating to the self).

Probably the most familiar set of explanations for ingroup bias are social-motivational. The common thread here is that group membership can validate or bolster the self, but only when the ingroup can be positively differentiated from contrasting outgroups [10]. In the context of the MGP, in which groups are largely devoid of meaning, individuals must engage in quite active fabrications if they are to meet this goal, potentially producing the range of biases we have discussed. But questions loom. First, where does the motivation to use groups as a tool of self-
enhancement come from in the first place? If it is itself learned then we must contend with the early emergence of minimal group effects (by age 3 and perhaps even earlier, well before we normally think of children as pursuing elaborated social identities [27,33,48]). Second, one might wonder why the extreme sparseness of minimal group settings makes individuals more desperate to bolster the ingroup rather than simply less interested in the whole affair. Indeed, naïve observers predict that bias will not appear in the MGP, suggesting that any such motivational process would have to operate outside of conscious awareness [49], and casting doubt on claims [50] that large effects in minimal group studies are the result of a demand effect or emerge only due to the absence of other usable cues. Third, if engaging in ingroup favoritism benefits the self, then we should be able to observe increases in self-esteem or related constructs following successful group enhancement. But this central prediction has not been generally supported [51]. Thus, while few would deny the relevance of social identity to intergroup cognition, it is not clear that it offers a persuasive explanation for the effects of mere membership.

A more recent suggestion is that social-motivational accounts get the causal arrow the wrong way around: It is not that we positively evaluate ingroups in order to positively value the self, but rather we positively evaluate ingroups because we already positively value the self [30,52,53]. Supporting this contention, implicit [30,52] and explicit [52,54] self-esteem are related to the magnitude of ingroup bias in the MGP. This view is also pleasingly parsimonious, linking ingroup bias to other phenomena such as general high self-regard [55], other ways in which things incidentally associated with the self come to be favored (e.g. the name-letter effect [56]; the endowment effect [57]), as well as broader theories of cognitive consistency [31]. Thus, ingroup bias as a generalization from or spillover of self-related positivity is a leading explanation for some membership effects.

Ingroup bias as invoked here is a broad phenomenon, a syndrome rather than a single effect. And some recent work begins to suggest that different aspects of this syndrome have distinct causal origins. In particular, the impact of mere membership on outcomes relating to coalitional reasoning and cooperative behavior may depend on quite different factors than those generating evaluative preferences. One piece of evidence in this regard comes from work suggesting that experimental manipulations explicitly ruling out ingroup interdependence (for
example by making it clear that the participant’s outcomes could not be affected by the decisions of other ingroup members) reduce or eliminate bias in cooperative behavior but not evaluative preferences [58,59], and indeed recent meta-analytic review concluded that manipulations of within-group interdependence have a large impact on cooperative behavior in both real and minimal groups [21]. An intriguing possibility emerging from this work (detailed in Box 3) is that there are (at least) two primary causal pathways leading to different classes of ingroup bias. A first evaluative form would be driven by self-related positivity spreading to the group, and could plausibly be considered a non-functional byproduct of representational structure such as a downstream consequence of tendencies towards high self-regard. A second coalitional form would be driven by subtle expectations about groups, perhaps in the form of an evolved coalitional reasoning system or a collection of tacit norms regarding ingroup loyalty and support [6,37]. A promising avenue for future work is the effort to test this dual pathway account.

Limits of mere membership

The consequences of mere membership are broad but not without limit, and the present perspective should not be taken as a call to disregard other critical factors. The clearest cases falling outside its scope are phenomena that rest on social learning. As we noted, evaluative preferences with respect to real groups tend to be stronger than those observed with minimal groups [47,60], and this creates an explanatory gap: what are the key ingredients that increase the impact of real-world groups? Familiarity is one likely contributor, as familiarity tends to produce liking in and of itself [61], and it is nearly by definition greater for real than minimal social groups. Thus, bias in favor of real groups could gradually increase as a function of increasing familiarity.

Stereotypes (learned associates of or inferences from category membership) are another critical area in which real and minimal groups will diverge. For example, some racial and ethnic stereotypes associate groups with highly specific positive and negative qualities (athleticism, math ability, shyness, aggression). Mere membership manifestly lacks the resources to explain these important phenomena. A related lacuna concerns effects that are derivable from group status, i.e. the position of a group within a broader social hierarchy. Critically, when a group is broadly stigmatized or otherwise devalued, members of that group tend to show weaker ingroup bias or
even a reversal to outgroup favoritism. For example, Black children in South Africa on average evaluate White South Africans more positively than they evaluate their own group [61,62]. Hierarchical social status of this kind is a property of broad social structures and must be inferred from environmental cues that go beyond membership [63-65]. Research comparing group biases in individuals who are and are not members of culturally advantaged or positively portrayed groups suggests that status and ingroup bias are largely additive, such that you can predict an individual’s degree of bias as the joint effect of status internalization (with higher status preferred) and ingroup bias (with the ingroup preferred) [66,67]. The dynamics of status internalization represents an area of inquiry that falls beyond the present account but which is surely be critical to understanding how representations of social groups develop in all their richness.

A more controversial case is outgroup homogeneity, the phenomenon in which outgroup members are perceived to be more similar to one another than ingroup members. This plausibly supports greater outgroup stereotyping because stereotypes are more likely to broadly apply to a homogenous group. A thorough but now dated review suggests that while robust in real groups, outgroup homogeneity does not appear in minimal groups [68]. However, this conclusion is in tension with more recent findings demonstrating preferential memory for and attention towards minimal ingroup members [46,69], which would seem to provide a causal pathway by which outgroup homogeneity, even if not present initially, could rapidly emerge.

This last case highlights an important subtlety in distinguishing effects of mere membership from those stemming from richer social learning: Even as they go beyond minimal group effects, social learning processes are themselves affected by learning biases that do follow directly from group membership. While this story can be told with adults, a developmental context makes the point most vividly. Beginning at least by the preschool years, membership in minimal groups biases the recall of valenced information about ingroups and outgroups [48], the interpretations of ambiguous intergroup events [70], the kind of information sought out and transmitted [71], and who one learns from or imitates [72]. Cumulatively this amounts to a learning gradient or canalization process in which ingroup-favoring information is preferentially encoded and transmitted. Thus, while things like stereotypes and status must be learned from social input, what is actually acquired will not be a veridical function of that input. Rather, it will
bear the stamp of biases that systematically support the internalization of ingroup-favoring group representations.

**Concluding Remarks: Towards a “Preparedness Account”**

Consider a young child’s first encounter with cues suggesting the presence of a previously unfamiliar social group. Given the dizzying range of human collectives, this new group may well seem mysterious. Indeed, a minimal group based on red versus blue shirts may not initially be any less opaque than, for example, a particular constellation of phenotypic cues that will eventually denote a racial distinction. Mastering the complex categorical terrain that encompasses a culture’s stock of group distinctions will require abstracting systematic variation from sparse cues, and it will pay to be charitable by assigning deep meaning to those cues [73,74]. The findings described here begin to illuminate the initial form of that charity: from the stark simplicity of “us and them” children and adults alike generate wide-ranging social preferences as well as sophisticated inferences about coalitional structure. Interpreted richly, these tendencies can be taken to reflect a deep and perhaps evolved preparedness for life in a coalitional world, a basic group template or set of innate moral norms that gets learning about groups off the ground. But it is important to note that even if one rejects this rich account in favor of leaner alternatives (such as the view that ingroup bias reflects a spillover from self-related positivity), the empirical phenomena summarized in Table 1 nonetheless loom large. Whether functional, incidental, or a bit of both, the fact remains: mere membership in a social collective is sufficient to explain a wide range of important intergroup phenomena, and clearly constitutes a rich suite of psychological processes powerfully channeling learning in the direction of ingroup favoritism. Understanding the nature of these processes and their downstream consequences must be a central goal of work in this area.

Deep questions remain (see Outstanding Questions). Chief among these is the effort to uncover which specific ingroup effects emerge via learning, as the consequence of evolved coalitional machinery, or as a result of some other cognitive process entirely (e.g. self-related positivity). Wherever these questions settle, the critical import of the work reviewed here is its implied demand for greater clarity about the causal underpinnings of ingroup bias, both in terms of the mechanisms that lead to its initial emergence and subsequent moderation, as well as the
extent to which observed phenomena arise from the general architecture of ingroup cognition versus specific facts about specific social targets. Progress can best be made by undertaking some crucial housekeeping, using the minimal groups procedure to place mechanisms and effects either inside or outside the diverse set of consequences that follow from mere membership in a social collective.
Box 1: The Most Famous Control Condition

A phenomenon as pervasive as ingroup bias admits of many candidate causes, and in the 1970s Henri Tajfel sought a rigorous paradigm for their systematic discovery [24]. His seemingly impeccable logic was to create groups so stripped of meaning as to be psychologically inert; such groups could then be the control condition over which likely causal factors could be carefully layered. Such is the humble origin of the minimal groups paradigm, which immediately yielded its startling surprise: despite these efforts, ingroup bias still sprung up. And it persisted, across early studies in which groups were based on novel dimensions such as (putative) art preferences [75] or tendencies to over- or under-estimate dot arrays [24], and through more strictly minimal groups based on explicit random assignment [30]. From the narrow range of dependent measures favored by Tajfel (resource allocation matrices), the scope of minimal group effects continued to expand until it extended to nearly wherever one looked (Table 1).

A clarifying early observation was that individuals thoroughly familiarized with the procedure and asked to predict what would occur within it robustly expect fair rather than biased behavior, silencing worries about potential demand effects (you can’t accede to a demand you don’t think is present) [49]. It also provides reason to think that the driving force producing these effects is what we might now call “implicit, in that participants are unaware of their own and others’ tendency to show bias.

To explain his surprising findings Tajfel initially appealed to a learned norm of group favoritism: through experience we have come to think we ought to benefit our ingroup members, perhaps because we expect them to benefit us [24]. Tajfel eventually rejected this account as more redescription than explanation [75], developing an alternative that came to form the heart of Social Identity Theory and its subsequent offshoots [10,11], in which bias emerges as a consequence of a tacit motivation to positively differentiate the ingroup from the outgroup to maintain or enhance self-esteem. Alternatively, rather than favoring groups in order to benefit the self, we may favor groups because we already favor the self, i.e. positive self-esteem as cause not consequence of ingroup bias [52]. While the dust has not yet settled on this debate (see main text and Box 3), the MGP remains perhaps the most famous and impactful control condition in the history of the behavioral sciences.
Box 2: Some examples of unexpected generality

The correct attribution of causation is a central goal of theory building. The range of consequences following from assignment to a minimal group suggests that in some cases, the cause of ingroup bias is as simple as membership in that group. When is membership a sufficient explanation, and when is it not? Widely shared social stereotypes are one case in which membership cannot be enough, as such content must be socially learned. But our intuitions are not always a good guide to whether a specific intergroup effect is or is not a stereotype. Take the tendency for angry facial affect to influence racial categorization, such that angry male faces are more likely to be categorized as Black than White (at least by Whites in North America) [44]. This seems a strong candidate for a learned stereotype, dovetailing neatly with our intuitions about cultural portrayals. A surprise, then, to find that angry faces are as likely to be over-categorized into a minimal outgroup as a racial outgroup [47]! What appeared to be an association between race and emotion is revealed to be a more general association between group and emotion. By contrast, this same study showed that the simple implicit preference for racial ingroups was considerably stronger than the preference for minimal ingroups, suggesting that the link between emotion and group membership is not simply another manifestation of an evaluative preference. More generally, while these findings are silent as to the potential impact of this or similar stereotypes in other contexts, it is clear that this particular effect should not be interpreted as a consequence of a specific learned stereotype.

A parallel story can be told with respect to the tendency to better remember faces of one’s ingroup, an effect most studied in the context of race and frequently attributed to the cumulative effects of experience interacting with ingroup members [45] leading to greater perceptual expertise, better individuation, and better recall. But an ingroup advantage in memory, and several related effects relating to decoding facial expressions, appear in very similar ways, sometimes even at indistinguishable magnitudes, with respect to minimal groups that do not differ in perceptual features or extent of prior exposure [46,69,76,77]. Further, encoding of racial ingroup faces is substantially degraded when the individuals are placed into crosscutting outgroups such as another university affiliation [79], though interestingly that same manipulation did not enhance encoding of racial outgroup faces. Taken together, these findings rule out purely perceptual explanations.
Such examples (and there are many) have a common structure: An effect previously attributed to a specific aspect of the intergroup terrain (learned stereotypes; acquired perceptual expertise) can now be reassigned or partially reduced to the general architecture of intergroup cognition. Racial outgroups are perceived as angry because they are outgroups, not because they are racial; and ingroup faces are remembered better because they are ingroups, not (at least not wholly) because they are more familiar.
Box 3: Dual Pathways to Different Membership Effects

The effects following from mere membership are varied and complex enough to suggest multiple causal pathways. An emerging possibility divides the terrain roughly in two. First are the primarily evaluative aspects of ingroup bias—those involving implicit or explicit links between the ingroup and positivity. These appear to be closely related to positive self-evaluations, such that when the self is linked to a particular group (e.g. via some explicit assignment), positivity associated with the self is transferred to the group, perhaps via an inferential or associative process seeking evaluative consistency [31]. Evidence for this view comes from consistent links between positive self-regard and these forms of ingroup bias [30,52,53], a relationship that is stronger in those more identified with the group [30,78].

The second constitutes a more abstract set of coalitional principles and expectations concerning group function. These encompass effects involving trust, coordination, cooperation, and loyalty, driven in some cases by a tacit norm of within-group reciprocity and support [58]. Supporting this distinction, experimental manipulations reducing interdependence between ingroup members (e.g. by explicitly highlighting the lack of opportunity for reciprocity) reduce or eliminate these forms of ingroup bias while leaving evaluative forms in place [79,80]. Relatedly, individual differences in group identification, which reliably predict evaluative preference, do not appear to predict bias in resource tasks [81], again suggesting distinct causal pathways.

Several additional considerations support this account. Meta-analysis of first-person evaluative bias in the minimal groups paradigm suggests that it is weaker than that generated by real groups [60], likely because of factors such as the increased familiarity and social identification engendered by real social identities (see main text). By contrast, meta-analysis of cooperative behavior [20] and behavioral discrimination [50] suggest that minimal group bias is equivalent or even stronger in those cases, a difference plausibly occurring because familiarity and social identification are not critical ingredients of these more general coalitional intuitions. Finally, there is the fact that ingroup preference per se is necessarily first-person, involving links between the individual and an ingroup, while coalitional reasoning is not restricted to first-party cases. Indeed, it is readily deployed to explain and predict third-party coalitional interactions beginning in infancy [34,36,38,82], phenomena difficult to account for via the mere extension of positivity to ingroup members.
Despite promise, this distinction may not be easy to maintain, in part because not all effects in Table 1 are easy to place within this taxonomy. Preferential learning, for example [62, 102], could plausibly be an outgrowth of preferences (copy those you like) or an outgrowth of a norm psychology designed to facilitate cooperation. Still, this framework for delineating lines of causal influence within the broad terrain of membership effects is a promising advance.
Table 1 (Key Table): Summarizing the Extent of Minimal Group Effects. Unless specified, effects are described in terms of a relative ingroup advantage and are provisionally sorted into evaluative, coalitional, learning, and miscellaneous categories. Representative (but not exhaustive) references are provided for each effect.

<table>
<thead>
<tr>
<th>Evaluative Effects</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive explicit attitudes</td>
<td>[30,72,83]</td>
</tr>
<tr>
<td>Positive implicit attitudes</td>
<td>[84-86]</td>
</tr>
<tr>
<td>Friendship preferences</td>
<td>[28,87]</td>
</tr>
<tr>
<td>Positive behavioral expectations</td>
<td>[88]</td>
</tr>
<tr>
<td>Positive trait attributions</td>
<td>[89,90]</td>
</tr>
<tr>
<td>Biased internalization of valenced actions</td>
<td>[91]</td>
</tr>
<tr>
<td>Positive evaluations of work products or related abilities</td>
<td>[90]</td>
</tr>
<tr>
<td>Relationships between explicit self-esteem and ingroup bias</td>
<td>[52,89]</td>
</tr>
<tr>
<td>Relationships between implicit self-esteem and ingroup bias</td>
<td>[30]</td>
</tr>
<tr>
<td>Cognitive-affective balance between bias, self-esteem, and identification</td>
<td>[30]</td>
</tr>
<tr>
<td>Happy as compared to angry faces more likely to be seen as ingroup</td>
<td>[47]</td>
</tr>
<tr>
<td>Greater empathy for pain</td>
<td>[92]</td>
</tr>
<tr>
<td>Greater emotional empathy</td>
<td>[93,94]</td>
</tr>
<tr>
<td>Favoritism in (hypothetical) allocation of positive and negative objects</td>
<td>[95,96]</td>
</tr>
<tr>
<td>Reverse-correlated ingroup faces more positive</td>
<td>[97]</td>
</tr>
<tr>
<td>More spontaneous positive trait inferences</td>
<td>[98]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coalitional Effects</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favoritism on allocation matrices</td>
<td>[24,41]</td>
</tr>
<tr>
<td>Favoritism on costly giving</td>
<td>[22,28,81]</td>
</tr>
<tr>
<td>Favoritism on non-costly giving</td>
<td>[29,72,87,89]</td>
</tr>
<tr>
<td>More punishment of outgroup aggressors</td>
<td>[99,100]</td>
</tr>
<tr>
<td>More attention to shared goals</td>
<td>[101]</td>
</tr>
<tr>
<td>Greater tendency to keep silent in the face of ingroup transgressions</td>
<td>[102]</td>
</tr>
<tr>
<td>Stronger tendency to keep secrets</td>
<td>[103]</td>
</tr>
<tr>
<td>Stronger reputational concerns</td>
<td>[81,104,105]</td>
</tr>
<tr>
<td>More cooperation in cooperative dilemmas</td>
<td>[80,106]</td>
</tr>
<tr>
<td>Greater trust</td>
<td>[87]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learning Effects</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>More trust in testimony</td>
<td>[72,107]</td>
</tr>
<tr>
<td>Bias in favor of consuming and transmitting positive information</td>
<td>[71]</td>
</tr>
<tr>
<td>Better face memory</td>
<td>[46,69,108]</td>
</tr>
<tr>
<td>Better memory for bias-consistent information</td>
<td>[48,88]</td>
</tr>
<tr>
<td>More positive interpretations of ambiguous events</td>
<td>[70]</td>
</tr>
</tbody>
</table>
More effective fear-conditioning to outgroup faces [109]
More imitation of ritual actions following ostracism [110]
Better identification of emotional expressions [76,77]
Outgroup males more readily associated with threat [109]
Differential impact of positive versus negative group information [91,111]

<table>
<thead>
<tr>
<th><strong>Miscellaneous Effects</strong></th>
<th><strong>References</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ingroups distinguished in N170 and other early ERPs</td>
<td>[112,113]</td>
</tr>
<tr>
<td>Differential activation in amygdala, FFA, mPFC</td>
<td>[114-117]</td>
</tr>
<tr>
<td>Error-related Negativity attenuated when observed by OG relative to IG</td>
<td>[118]</td>
</tr>
<tr>
<td>More spontaneous representation of ingroup actions</td>
<td>[119]</td>
</tr>
<tr>
<td>Hand movements judged faster</td>
<td>[120]</td>
</tr>
<tr>
<td>Self-related traits projected to ingroup</td>
<td>[121,122]</td>
</tr>
<tr>
<td>Lower threshold for mind perception in human-doll morphs</td>
<td>[19]</td>
</tr>
<tr>
<td>Ingroup members judged more similar</td>
<td>[27-29,90]</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to thank Jay Van Bavel, Paul Bloom, and Jan Engelmann for helpful comments on an earlier version of this manuscript, and Joshua Knobe and the members of the Social Cognitive Development Lab at Yale for helpful discussions. I also gratefully acknowledge the support of the John Templeton Foundation (Award 56036).
References


Patterson, M.M. and Bigler, R.S. (2016) Effects of consistency between self and in-group on children’s views of self, groups, and abilities. Social Development 43, 55–18


Navarrete, C.D. et al. (2012) Fear is readily associated with an out-group face in a minimal group context. *Evol Hum Behav* 33, 590–593


Molenberghs, P. et al. (2013) Seeing is believing: Neural mechanisms of action–perception are biased by team membership. *Hum Brain Mapp* 34, 2055–2068
