RUNNING HEAD: From Categories to Exemplars

From Categories to Exemplars (and Back Again)

Yarrow Dunham

University of California, Merced

Juliane Degner

University of Amsterdam

Corresponding Author:

Yarrow Dunham

ydunham@ucmerced.edu

To appear in:

S.A. Gelman and M.R. Banaji (Ed.), *Social Cognitive Development*. Oxford University Press.

From Categories to Exemplars (and Back Again)

The psychological study of the development of prejudice is now some 80 years old. At this point, perhaps its least controversial conclusion is that prejudice emerges quite early, certainly by age four or five in the context of race in North American settings (e.g. Aboud, 1988). The interest of these findings—and the surprise they often elicit in educators and parents—stems from the assumption that *prejudice* (the affective or attitudinal dimension of intergroup bias) produces *discrimination* (the behavioral dimension). Put differently, attitudes are assumed to be causal players, the "under the hood" psychological entity that drives behavior (e.g. Allport, 1935). And certainly it is generally the case that those with more negative intergroup attitudes do discriminate more (e.g. Ajzen & Fishbein, 1977; Greenwald, Poehlman, Uhlmann, & Banaji, 2009).

But what do we know about how prejudice and discrimination relate to one another in childhood? When assessed with traditional verbal measures, children between the ages of 5 and 7 show the strongest prejudice observed across the lifespan (a claim recently validated meta-analytically: Raabe & Beelmann, in press). Thus, we have every reason to predict that children in this age range would engage in highly discriminatory behavior, indeed, in the *most* discriminatory behavior that we observe across the lifespan. Do they? While the question of whether children in the preschool years discriminate *at all* is still somewhat open, it seems clear that they do not discriminate *a lot*. Studies tend to find at most low levels of discrimination, and further suggest that what is observed cannot be securely attributed to race. That is, the importance of race is reduced when other, most notably socio-economic factors, are statistically controlled for (Graham et al, 1998; Kupersmidt, DeRosier, & Patterson, 1995; Singleton & Asher, 1979). Thus, the cautious conclusion is that while children do care about some things that are *correlated* with race (for example, cues associated with poverty), at least in the preschool to early elementary school ages, *race itself* exerts little influence on their behavioral tendencies in the real ecology of a playground or classroom.

How troubling should we consider this disconnect between attitudes and behavior? Certainly if we survey the field we will find that failures to find attitude-behavior correspondence are nearly as

common as successes, an observation that has produced its share of hand-wringing (e.g. Wicker, 1969). In response, the field has generated some convincing reasons for the apparent disconnect, as well as some strategies to overcome it. Most prominently in the domain of prejudice, attitudes and behavior might not line up if individuals are reluctant to report on their true attitudes, for example because admissions of prejudice carry social costs. But of course younger children *do* express prejudice at extremely high rates (for example, expressing preference for their racial ingroup on about 80% of trials Dunham, Baron, & Banaji, 2006), making it implausible that they are greatly influenced by social norms against its expression (which is not to say they are never influenced by such norms; see Rutland, Cameron, Milne, & McGeorgoe, 2005). Relatedly, social psychologists have suggested that even those who consciously hold little prejudice may have (if only inadvertently) internalized negative representations of outgroups, and that it is this "implicit" form of attitude that drives much discriminatory behavior (e.g. Fazio, Sanbonmatsu, Powell, & Kardes, 1986). If this contention is correct, it raises the possibility that children tend not to discriminate because they have not yet acquired implicit biases. But this possibility also fails to pass empirical muster: certainly by early elementary school, children show implicit biases every bit as strong as their adult counterparts (reviewed in Dunham, Baron, & Banaji, 2008).

Thus, children show strong prejudice at both the explicit and implicit level, but these attitudes do not reliably manifest themselves in behavior. Given that the usual avenues towards closing this gap (developed with adults in mind) do not seem workable here, we must look elsewhere. Why not begin, then, with careful consideration of the constructs themselves? Stereotypes and prejudices are at their core *category-level* phenomena. That is, they say nothing about individuals-*qua-individuals*, only about individuals-*qua-exemplars* of specific categories. This is just to say that stereotypes and prejudices are about groups and their members (e.g., semantic associations between groups and traits; Greenwald et al, 2002). Perhaps this seems obvious, but we belabor the point because it brings into focus an important gap between the attitude and its behavioral manifestation, a gap that may be particularly crucial in early childhood. More precisely, consider what is required for a group-level representation to impact a given interpersonal interaction (and in so doing produce discrimination).

A potential instance of discrimination is an *interpersonal interaction*, and as such is first and foremost an interaction with an *individual*. Take it as a given that this individual falls under the scope of a negatively evaluated category. But that category-level evaluation is not *inherent* in the individual; to do any work, it must actually be activated in the course of the interaction. And here is the gap that we mean to highlight. We suggest that children acquire negative evaluations of social categories (qua categories) well before those categories are routinely brought to bear in the course of social interaction. In a sense, the prejudice floats above the fray, exerting causal influence only when the context makes the categories particularly salient, leading to their activation and subsequent application.

By way of example, consider the familiar case of La Pierre (1934), who found that restaurant and hotel proprietors in the early 1930s overwhelmingly rejected a written entreaty to accept a Chinese couple as customers, but even more overwhelmingly accepted them without visible malice or even hesitancy when they actually visited unannounced. While certainly many factors could make responses to a letter and to an actual encounter diverge, we focus here on just one, concerning the extent to which the category is operative in the potential discriminator's mind. The in-person encounter is with a couple that is, among other things, neatly dressed, kind looking, elderly, Chinese. Whether, in the midst of these factors, the couple's 'Chineseness' is activated, and if so whether it is strong enough to override the multitude of other factors, become the operative questions. By contrast, a written entreaty to accept "a Chinese couple" is explicitly at the level of a (named, stigmatized) category. The reader cannot but have their concept of "Chinese" activated, and now it will directly involve itself in their behavioral response.

Thus, negative attitudes towards one category are but one force among many, and they will not necessarily win out when pitted against other factors, such as additional attitudes or contextual cues. Surely that is true: even at its best, the link between an attitude and a behavior will beappear in some cases because it is but one force and that will often be overpowered by others (though this is no doubt true). Rather, we are also pointing out that a category-level representation can only affect an individual when that individual is in fact actually placed into the category in the moment of the interaction. The

category must be active to exert its influence at all, to even be one of the causal forces careening through the cognitive calculus that leads to a behavior.

Even in adults, category activation is no sure thing. For example, while there is evidence that in some cases both adults and elementary school children automatically categorize race and gender (as shown, for example, by memory confusion paradigms: Bennett & Sani, 2003; Taylor, Fiske, Etcoff, & Ruderman, 1978), it is by no means the case that such categorization is inevitable. For example, when placed under cognitive load, adults failed to encode the race of a woman they viewed on a video screen, as evidenced by the lack of stereotype-based intrusions into a subsequent word completion task, suggesting that category application requires mental resources (Gilbert & Hixon, 1991), even in adults, who have learned and probably over-learned racial categories. And this gives us every reason to think that category application will be even less sure in children, for whom racial categories are not yet firmly established. In a recent study, we found that 5-yr-olds could accurately apply racial categories in a forced choice paradigm only about 65% of the time—a figure that just exceeded chance performance, but which speaks of a surprisingly fragile mapping between category labels and perceptual features (despite stimuli that, through their use of highly representative racial exemplars, likely make the category boundary artificially clear).

Thus, one reason a category might not be deployed by children is that it is still fragile, poorly integrated with diagnostic perceptual cues. But even once children *can* categorize with a high degree of accuracy, there remain further reasons why children might be less likely to apply it. Consider the range of social categories adults can apply with a high degree of accuracy. Amongst these, they only *automatically* deploy a limited set (e.g. race and gender, and even those not in all cases, as we discussed above). This suggests that only categories that are deeply ingrained and culturally reinforced will be brought to bear habitually, with others requiring more effort and/or more contextual cues to initiate their deployment. Children, then, might go through a period of having racial categories but not deploying them unless the context provides strong cues suggesting they are relevant.

Our consideration of this question has led us to suspect that many of the most common assessments of prejudice provide just this sort of cue; that is, they strongly encourage activation of racial categories. Take for example the family of methods derived from the Clark Doll Task (Clark & Clark, 1939), in which contrasting 'minimal pairs' of individuals are presented, and children are asked to say who they like better, would prefer to play with, etc, in a forced-choice manner. The pairs are designed to be as closely matched as possible on dimensions other than race (e.g. typically only same gender pairs are employed, the faces are matched on age, attractiveness, and clothing, and when drawings are used only a few prototypical cues such as skin color or hair type vary), and multiple trials are presented sequentially. The sequential presentation of minimal pairs that differ most prominently (or even exclusively) along a racial dimension is nothing if not a powerful way to make that dimension salient! That is, the methodology strongly encourages the activation of racial categories and their attendant evaluations—that is, a category-based stereotype or prejudice. Imagine contrasting prejudice assessed through this 'minimal pairs' methodology with a task in which individuals are presented one at a time, and ratings of each individual are made on a continuous dimension of liking that is independent of ratings of other targets. Here, while race varies across individuals, other factors do as well, like gender, attractiveness, clothing, hair style, etc. While care is needed to ensure that across the entire stimulus set attractiveness and other factors are equated, this task differs in that the child is free to direct their attention towards whatever factors are most naturally salient to him or her. To the extent that race is among them, it should affect preferences, but if race is impactful only when the context makes it salient, we should see considerably weaker preferences on such a task.

We recently explored this possibility in 5 and 7-yr-old children. As expected, when presented with a minimal pairs methodology, both 5 and 7-yr-olds robustly preferred the White ingroup, choosing the same race peer almost 70% of the time. However, in a single-target rating task, 5-yr-olds showed no evidence whatsoever of in-race preferences, while 7-yr-olds showed a statistically significant but still weak preference for their racial ingroup. That is, when individuals are presented *qua individuals*, freely varying along a number of dimensions, the ubiquitous finding of in-race preference in young children

disappears! At risk of repetition: When the situation does not make a racial category salient, the category-level evaluation will not necessarily be active, and so will not influence how an individual is evaluated.

What about when attitudes are assessed at the implicit level? As with adults, the most widely used measure of implicit attitudes in children is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwarz, 1998). Multiple investigators have now found that, when measured with the IAT, racial preferences appear in adult-like forms from as early as they can reliably be measured (for a review, see Olson & Dunham, 2010). But the IAT is first and foremost a category-based preference measure, in that it involves the explicit categorization of individuals into named groups (e.g., faces are explicitly classified by race). This is by no means a weakness of the measure, which was specifically designed to measure category-based evaluations. However, it cannot thereby answer the question of whether that negative category-level evaluation affects how an individual is evaluated *in vivo* (again, remember the instructive case of La Pierre's Chinese couple). One can contrast the IAT with methods such as evaluative priming, in which faces of individuals (varying in race) precede words or pictures which must be categorized by valence, and the question becomes whether faces of a certain race facilitate responding, e.g. whether Black faces facilitate responding to negative targets. Note that because no explicit categorization is involved, we would expect an influence of racial category only if the participant spontaneously brings that category to bear. And indeed, when assessed in this manner, participants appear 'less prejudiced', at least when we consider the proportion of a participant pool who manifest pro-White, anti-Black implicit bias in the US—unless some other aspect of the procedure makes racial categories salient (for example, being told you will be asked to recall how many individuals fell into each racial category), in which case both families of measures produce similar rates of bias (Olson & Fazio, 2003).

Bringing it back to development, the line we are pursuing here suggests that while category-based assessments will reveal early-emerging implicit preference, assessments that measure the evaluation of *individuals* (who happen to be members of a given category) will only reveal race-based implicit preferences later in development. This is exactly what was found in a study of White German and Dutch children's implicit attitudes towards Turkish immigrants (Degner & Wentura, 2010). While children

between 9 and 15 all showed a uniform degree of pro-White bias on the IAT, when measured in an evaluative priming paradigm there was a linear increase in the strength of preference with age, with the youngest children showing no evidence of ingroup preference at all. Of course, measures differ in many ways, but to more accurately peg the difference to explicit use of categories, Degner and Wentura followed the lead of Olson and Fazio (2003) described above by asking children to categorize each prime by race immediately after responding to the target picture that followed it. With this manipulation, their younger participants now showed robust implicit bias on the priming measure as well, suggesting that just like in adults, category activation is the key factor driving the presence of implicit bias, but that 'baseline' category activation levels are lower in children than in adults.

Of course, caution must be taken when comparing these results to the more familiar case of race in America. Could these results depend on the European context, in which Turkish immigrants might be both more perceptually similar (to children's white majority ingroup) and less familiar than African-Americans in the US? To answer this question, in a recent study we measured White elementary school children's implicit race preferences towards African-Americans using both the IAT and the Affect Misattribution Procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2004), a priming-based methodology that depends for positive results on different stimuli arousing different degrees of positive and negative affect, but which does not include any explicit categorization. Replicating prior work, both 6 and 9-yr-old children showed robust implicit preferences for White over Black when measured with the IAT. However, only 9-yr-olds showed implicit White over Black preference when measured with the AMP. Thus, while when evaluating the categories themselves, 6-yr-olds show a pro-White implicit preference, but when exposed to faces of Black and White children, these same children's responses do not appear to be affected by race.

We interpret this pattern of results as showing that children acquire a *category-level* evaluation of named social categories such as race well before they make habitual use of that category in the course of a routine social encounter (while they do not draw out the implications in the same way, related suggestions can be found in Hirschfeld, 1996 and Quintana, 1998). This means that in most such encounters, the

causal pathway by which the attitude could influence behavior is not present—either because the child has not yet learned to apply the category accurately, or because, despite having learned how to apply it, they do not yet do so automatically. In either of these cases, the category-level evaluation sits idly by, doing no work at all, because the interlocutor has not actually been categorized into the relevant group. Of course, many questions remain. Most prominently, what leads to the eventual automatization of category application? Is it simply a case of more experience, or are their specific forms of cultural input that lead to this outcome, and explain why some but not all categories become so automatic? Documenting the path to automaticity, and the cultural inputs that support it, is therefore a primary future research goal.

Another critical piece involves actually doing the work of assessing how different forms of attitude affect children's behavior. Our claim entails the strong prediction that, at least in younger children, exemplar-level measures like priming will be better predictors of discrimination than will category-level measures like the IAT, because they measure the spontaneous application of the racial categories. Despite the practical challenges of measuring discriminatory behavior in children, this central prediction of our proposal needs to be investigated.

Our main contribution is to note that an attitude towards a group is not the same thing as an attitude towards an individual who happens to be a member of that group. To equate those two is to presume an additional cognitive step, the step from category *possession* to category *application*. We believe this step is a later developmental 'accomplishment', and its absence in early childhood may well explain the apparent disconnect between attitudes and behavior in young children. This conceptual point has a practical sibling: researchers must carefully consider whether the methods they choose provide leverage on the research questions they are pursuing. Some measurement techniques necessarily implicate categories and their attendant evaluations, or provide a strong contextual push in that direction. Others include tacit assumptions about whether or not these same categories are active during tasks of varying subtlety, or seek to measure that activation directly. These differences in methods parallel differences in substantive theoretical and empirical claims regarding how categories become guides to behavior. We

hope researchers will keep the distinctions firmly in mind, and choose methods that match their substantive research goals.

References

Aboud, F. E. (1988). *Children and prejudice*. Oxford: Basil Blackwell.

Ajzen, I., & Fishbein, M. (1977). Attitude–behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84,* 888–918.

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *A Handbook of Social Psychology*. (pp. 789–994). Worcester, MA: Clark University Press.

Bennett, M. & Sani, F. (2003). The role of target gender and race in children's encoding of category-neutral person information. *British Journal of Developmental Psychology, 21*, 99-112.

Clark, K.B. & Clark, M.K. (1939). The development of consciousness of self and the emergence of racial identification in negro preschool children. *Psychological Bulletin, 10*, 591-599

Degner, J. & Wentura, D. (2010). Automatic Prejudice in Childhood and Early Adolescence. *Journal of Personality and Social Psychology. 98,* 356-374.

Dunham, Y., Baron, A. S., & Banaji, M. R. (2006). From American city to Japanese village: A cross-cultural investigation of implicit race attitudes. Child Development, 77, 1268 – 1281.

Dunham, Y., Baron, A.S., & Banaji, M.R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences, 12*(7), 248-253.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C. & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229-238.

Gilbert, D.T. & Hixon, H.J. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology 60*(4), 509-517.

Graham, J.A., Cohen, R., Zbikowski, S.M., & Secrist, M.E. (1998). A longitudinal investigation of race and sex as factors in children's classroom friendship choices. *Child Study Journal, 28*(4), 245-267.

Greenwald, A. G., McGhee, *D*. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17–41.

Hirschfeld, L.A. (1996). *Race in the Making: Cognition, Culture, and the Child's Construction of Human Kinds.* Cambridge: MIT Press.

Raabe, T. & Beelmann, A. (in press). Development of Ethnic, Racial, and National Prejudice in Childhood and Adolescence: A Multinational Meta-Analysis of Age Differences. *Child Development.*

Kupersmidt, J.B., DeRosier, M.E., & Patterson, C.P. (1995). *Journal of Social and Perosnal Relationships, 12*(3), 439-452.

LaPiere, R. T. (1934). Attitudes vs. Actions. Social Forces, 13(2), 230-237.

Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development, 76*(2), 451 – 466.

Olson, K.R. & Dunham, Y.D. (2010). The development of implicit social cognition. In B. Gawronski & B. Keith Payne (Eds). *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications*. New York: Guilford.

Olson, M.A. & Fazio. R.H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science, 14*(6), 636-639.

Quintana, S. M. (1998). Children's developmental understanding of ethnicity and race. *Applied and Preventive Psychology, 7,* 27–45.

Singleton, L.C. & Asher, S.R. (1979). Racial integration and children's peer preferences: An investigation of developmental and cohort differences. *Child Development, 50*, 936-941.

Taylor, S.E., Fiske, S.T., Etcoff, N.L., & Ruderman, A.J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology 36*(7), 778-793.

Wicker, A. W. (1969). Attitudes versus actions: The relationship between verbal and overt behavioral responses to attitude objects. *Journal of Social Issues, 25*, 41-78.