Can race be erased? Coalitional computation and social categorization

Robert Kurzban†, John Tooby, and Leda Cosmides

Center for Evolutionary Psychology, University of California, Santa Barbara, CA 93106-2070

Communicated by Roger N. Shepard, Stanford University, Stanford, CA, October 12, 2001 (received for review May 28, 2001)

Previous studies have established that people encode the race of each individual they encounter, and do so via computational processes that appear to be both automatic and mandatory. If true, this conclusion would be important, because categorizing others by their race is a precondition for treating them differently according to race. Here we report experiments, using unobtrusive measures, showing that categorizing individuals by race is not inevitable, and supporting an alternative hypothesis: that encoding by race is instead a reversible byproduct of cognitive machinery that evolved to detect coalitional alliances. The results show that subjects encode coalitional affiliations as a normal part of person representation. More importantly, when cues of coalitional affiliation no longer track or correspond to race, subjects markedly reduce the extent to which they categorize others by race, and indeed may cease doing so entirely. Despite a lifetime's experience of race as a predictor of social alliance, less than 4 min of exposure to an alternate social world was enough to deflate the tendency to categorize by race. These results suggest that racism may be a volatile and eradicable construct that persists only so long as it is actively maintained through being linked to parallel systems of social alliance.

Throughout our species' history, intergroup conflict depended on the categorization of the social world into *us* versus *them*. When this divide occurs along racial lines, this categorization and its malignant consequences appear capable of persisting stably. Indeed, ingroup favoritism paired with outgroup indifference or hostility appears to exist in all human cultures (1, 2). The simple act of categorizing individuals into two social groups predisposes humans to discriminate in favor of their ingroup and against the outgroup in both allocation of resources and evaluation of conduct (2–7). Following on historical experience, field and laboratory studies have confirmed that this behavior is remarkably easy to elicit: people discriminate against outgroups even when they are assigned to groups temporarily and anonymously by an experimenter who uses dimensions that are trivial, previously without social significance, and random with respect to any real characteristics of the individuals assigned (2–8). Given that categorizing people into groups along nearly any dimension elicits discrimination, it would be discouraging to learn that the human mind was designed such that people cannot help categorizing others by their race. This would imply that racism is intractable.

Yet it has been claimed, with considerable empirical support, that encountering a new individual activates three "primitive" or "primary" (9–12) dimensions—race, sex, and age—which the mind encodes in an automatic and mandatory fashion (i.e., across all social contexts and with equal strength; refs. 10–15). These dimensions can be encoded without other individuating information; e.g., one might recall that one's new neighbor is a young, white woman, without remembering anything else about her (11–15). Over the last two decades, considerable effort has been expended on the search to find conditions under which race is not encoded, so far without success (14, 15).

Despite the evidence in favor of these claims, an evolutionary analysis indicates that one of them is likely to be wrong. Although selection would plausibly have favored neurocomputational machinery that automatically encodes an individual's sex and age, "race" is a very implausible candidate for a conceptual primitive to have been built into our evolved cognitive machinery. During our evolutionary history, our ancestors would have inhabited a social world in which registering the sex and life-history stage of an individual would have enabled a large variety of useful probabilistic inferences about that individual. In contrast, ancestral hunter-gatherers traveled primarily by foot and, consequently, residential moves of greater than 40 miles would have been rare (16). Given the breeding structure inherent in such a world, the typical individual would almost never have encountered people sampled from populations genetically distant enough to qualify as belonging to a different "race" (even assuming that such a term is applicable to a nonpolytypic species such as humans, in which the overwhelming preponderance of genetic variation is within population and not between population, and at most geographically graded rather than sharply bounded) (17, 18). If individuals typically would not have encountered members of other races, then there could have been no selection for cognitive adaptations designed to preferentially encode such a dimension, much less encode it in an automatic and mandatory fashion.

Accordingly, we propose that no part of the human cognitive architecture is designed specifically to encode race. We hypothesize that the (apparently) automatic and mandatory encoding of race is instead a byproduct of adaptations that evolved for an alternative function that was a regular part of the lives of our foraging ancestors: detecting coalitions and alliances. Hunter-gatherers lived in bands, and neighboring bands frequently came into conflict with one another (19–21). Similarly, there were coalitions and alliances within bands (22), a pattern found in related primate species and likely to be far more ancient than the hominid lineage itself (23). To negotiate their social world successfully, and to anticipate the likely social consequences of alternative courses of action, our ancestors would have benefited by being equipped with neurocognitive machinery that tracked these shifting alliances. Computational machinery designed to detect coalitions and alliances in the ancestral world, if well designed, should be sensitive to two factors: (*i*) patterns of coordinated action, cooperation, and competition, and (*ii*) cues that predict—either purposefully or incidentally—each individual's political allegiances (24–26). Like other behaviors, actions that reveal coalitional dispositions are usually transitory, and so are frequently unavailable for inspection by others when decisions relevant to coalitional affiliation need to be made. Accordingly, alliance-tracking machinery should be designed to note these rare revelatory behaviors when they occur, and then use them to isolate further cues that happen to correlate with coalition but that are more continuously present and perceptually easier to assay. Such cue mapping allows one to use the behavior of some individuals to predict what others are likely to do. Because this circuitry detects correspondences between allegiance and appearance, stable dimensions of shared appear-

†To whom reprint requests should be addressed. E-mail: rkurzban@ucla.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

ance—which may be otherwise meaningless—emerge in the cognitive system as markers of social categories. Coalitional computation increases their subsequent perceptual salience and encodes them at higher rates. Any readily observable feature—however arbitrary—can acquire social significance and cognitive efficacy when it validly cues patterns of alliance. Ethnographically well-known examples include dress, dialect, manner, gait, family resemblance, and ethnic and coalitional badges.

Adaptations for coalition detection are hypothesized to resemble adaptations designed to encode sex and life-history stage in some respects, but not others. Ancestrally, an individual's sex was fixed for life, but his or her coalitional membership was not. New patterns of alliance typically emerge whenever new issues arise whose possible resolutions differentially impact new subsets of the local social world. Consequently, coalitions changed over time and varied in internal cohesion, duration, and surface cues. To track these changes, cue validities would need to be computed and revised dynamically: no single coalitional cue (including cues to race) should be uniformly encoded across all contexts. Furthermore, arbitrary cues—such as skin color—should pick up significance only insofar as they acquire predictive validity for coalitional membership (26).

In societies that are not completely racially integrated, shared appearance—a highly visible and always present cue—may be correlated with patterns of association, cooperation, and competition (26). Under these conditions, coalition detectors may perceive (or misperceive) race-based social alliances, and the mind will map race onto the cognitive variable *coalition*. According to this hypothesis, race encoding is not automatic and mandatory. It appeared that way only because the relevant research was conducted in certain social environments where the construct of race happened, for historical reasons (27), to be one valid probabilistic cue to a different underlying variable—one that the mind was designed to automatically seek out—coalitional affiliation (24–26).

Predictions.

Because coalitional allegiance is hypothesized to be the actual cognitive variable that the mind is encoding when it is registering race, we predict the following.

Prediction 1. Race *per se* will not, after all, have the properties previously ascribed to it, i.e., it will not be encoded across all social contexts and with equal strength. Manipulating coalitional variables should introduce the long sought-after variability in its encoding.

Prediction 2. Shared visual appearance is not necessary for coalition encoding. People will encode coalitional alliances as part of person representation even when there is no similarity of appearance within a coalition.

That is, coalitional categories are not built on a foundation of shared appearance and can readily exist without it. Shared appearance will acquire (or lose) significance only to the extent that it provides a cue to this underlying political structure.

However, a correlation between shared appearance and behavioral cues of alliance can reinforce and stabilize a coalitional categorization, leading to a counterintuitive prediction.

Prediction 3. Arbitrary cues other than racial appearance can be endowed with the same properties that race has previously exhibited by linking them to coalition membership.

Even more significantly, if race is merely a proxy for coalition—a cue used to infer a person's alliances—then it is predicted that:

Prediction 4. The strength of race encoding will be diminished by creating a social context in which (*i*) race is no longer a valid cue to coalition, and (*ii*) there are alternative cues that do reliably indicate coalitional affiliation.

There are no *a priori* grounds for predicting exactly how fast the human mind should be able to discard ontogenetically long-standing coalitional categorizations in favor of novel ones. But if our cognitive adaptations are well engineered by selection, then they might be able to detectably alter such categorizations even within the brief periods available during an experiment.

Finally, because sex is hypothesized to be a true and stable primary dimension of person representation similar to but independent of age and coalition, we predict that:

Prediction 5. Sex will be encoded far more strongly than race, even in contexts in which it is irrelevant to coalition and task.

Prediction 6. The encoding of sex, an independent dimension, will not diminish coalition encoding, even though sex will be encoded far more strongly than race.

The impact of the encoding of sex on coalition is introduced to allow the evaluation of a plausible counterhypothesis: that attention is zero sum, and so a negative relationship between any two dimensions of encoding shows only that attention is finite. Demonstrating that sex encoding does not diminish coalition encoding eliminates attentional constraints as an alternative explanation for any race reduction effect (prediction 4). If coalition is encoded just as strongly when the competing dimension is sex—which is encoded more strongly than race—then no domain-general capacity limit is preventing race from being encoded as strongly as sex. If no ceiling on a domain-general attentional system has been reached, then attentional constraints cannot explain any reduction in race encoding observed in this paradigm. Eliminating this alternative would reinforce the interpretation that race encoding is diminished because race is merely a proxy for the deeper mental category, *coalition*—and not for some incidental reason.

We present results supporting these six predictions below.

Methods

To assess encoding, we used a method standard in the literature, the memory confusion protocol developed by Taylor *et al.* (13). It uses errors in recall to unobtrusively reveal whether subjects are categorizing target individuals into groups and, if so, what dimensions they are using to do so (see *Appendix*). Subjects are asked to form impressions of individuals whom they will see engaged in a conversation. They then see a sequence of sentences, each of which is paired with a photo of the individual who said it. Afterward, there is a surprise recall task: the sentences appear in random order, and subjects must attribute each to the correct individual. Misattributions reveal encoding: subjects more readily confuse individuals whom they have encoded as members of the same category than those whom they have categorized as members of different categories. For example, a citizen of Verona who had encoded coalition membership would make more within-category errors—errors in which s/he confused, say, a Capulet with a Capulet (or a Montague with a Montague)—than between-category errors—ones in which s/he confused a Capulet with a Montague or vice versa (this relationship will hold for data that are corrected to equalize base rates; see *Appendix*).

Experiments 1 and 2 were designed to test predictions 1–4 above. The first step was to test the prediction that subjects' minds could be caused to treat the dimension *coalition* in a fashion parallel to sex, age, and especially race, i.e., to see whether subjects would encode *coalition* incidentally as part of impression formation, without having been instructed to do so (predictions 2 and 3). A widely held counterhypothesis is that similarity in appearance is the origin and foundation of social categorizations such as sex, age, and race, which only then acquire social significance. Because our view is that coalition encoding can be elicited even in the absence of shared visual appearance (prediction 2), there were no visual cues to coalition membership in experiment 1. Each speaker (represented by a photo) was a young man, and all were identically dressed. The only way to infer the coalitional allegiances of these individuals was from the content and sequence of their utterances: speech

acts implying self-inclusion, allegiance to, exclusion from, or enmity toward one or another of two contending groups (hereafter, *verbal allegiance cues*—e.g., "you were the ones that started the fight"). From these cues, four of the eight speakers could be categorized as belonging to one coalition, and four to the other.

Thus, in experiment 1, nothing in the appearance of the individuals offered the potential for subjects to correctly sort them into two coalitions (a fact logically guaranteed by the experimental design: for each subject, the computer randomly generated a new assignment of photographs to coalition). This provided a very stringent test of coalition encoding, methodologically adverse to our hypothesis. Categories such as race and sex are on display in the appearance of the individual both at the time of encoding and during the recall task (facilitating the use and acquisition of these two categories). But coalitional membership was not determinable from appearance at any time during experiment 1, and no cues to coalition membership—visual or verbal—were present during the recall task. These handicaps would necessarily weaken the detectability of coalitional categorization, compared with race and sex, making its spontaneous extraction and use, if found, all of the more remarkable.

According to our central hypothesis, two factors—patterns of alliance and shared appearance—conspire to reinforce and stabilize a coalitional categorization along racial lines. The same theory implies, however, that other coalitional categorizations will be encoded in a race-like manner if the second factor—shared appearance—is present and tracks behavioral cues of coalitional alliance (prediction 3). This should be true even when shared appearance is created by using a visual cue that is arbitrary and impermanent. Prior results with the memory confusion protocol show that differences in shirt color are not spontaneously encoded when they have no social significance (15), so this was a good candidate for an arbitrary cue. Thus, experiment 2 was identical to experiment 1, except we added this arbitrary feature of shared appearance to the coalitions that we had manufactured through verbal allegiance cues. We used image manipulation software to give the individuals either gray or yellow shirts; individuals whose verbal statements implied allegiance to the same coalition wore the same color. Thus, subjects in experiment 2 could infer coalition membership from verbal allegiance cues, from cues of shared appearance (shirt color), or both, because they tracked each other. If prediction 3 is correct, then coalition will be spontaneously encoded in experiment 2 just as strongly–or more so–than race is. (Strength of encoding is indexed by effect size, *r*; ref. 28.)

An equally important goal shaping the design of these experiments was to observe how the creation of an independent system of coalitional categorization would affect racial encoding (predictions 1 and 4). The experimental context implied a conflict between two rival coalitions, but these were constructed such that race was uncorrelated with coalition membership: the rival four-person coalitions were each composed of two Euro-American men and two African-American men. This design allowed us first (experiment 1) to replicate the basic phenomenon of racial encoding and, second (experiment 2), to see whether the tendency to encode race could in fact be reduced by the introduction of a visually accessible cue to a nonracial coalitional division (thus testing prediction 4). If the human brain contains neurocomputational machinery for tracking coalitional alliances, then constructing a new social environment in which coalition is uncorrelated with race should weaken the preexisting weight given to race as a cue to coalition within that context. The more obvious and relevant the alternative coalitional division, the less race should be encoded. But if prior claims are correct—if race is a primary dimension of person representation, encoded with equal strength across all social contexts—then constructing this new social environment should make no difference. Indeed, if race is a primary dimension and coalition is not, one might expect race to overwhelm coalition encoding across contexts.

Even if our hypothesis were correct, for the effects to be observable, the machinery would have to be very efficient at updating and tracking even short-term changes in alliance, dynamically recomputing cue validities associated with newly relevant coalitions, even with very limited exposure. Otherwise, a few minutes in an experiment in which race was not predictive of coalition could not detectably impact the accumulated effect of years of exposure to a world in which race mattered.

Results

The results of experiment 1 were as follows:

(*i*) Even though race existed as a competing and visible dimension, subjects did indeed encode a new dimension, coalition membership, doing so solely on the basis of verbal cues of implied affiliation (confirming prediction 2): they made more within than between coalition errors [$n = 55$, mean difference, $M = 0.90$ (SD = 2.76), $t^*(54) = 2.41$, $P = 0.0096$; see *Appendix*, Note 1].

(*ii*) Consistent with previous findings, subjects also encoded information about the race of targets, forming social categories on this basis as well [$M = 2.30$ (SD = 2.76), t^* (54) = 6.71, $P < 6.2 \times 10^{-9}$].

(*iii*) In experiment 1, where verbal allegiance cues were the only basis for inferring coalition membership, and coalition was visually undetectable, the effect of race was twice as large as the effect of coalition (effect size, *r*, for coalition versus race: 0.31 and 0.67, respectively).

Experiment 2 was identical to experiment 1, with one exception: cues to coalitional affiliation were amplified by giving each coalition its own uniform color, either gray or yellow, providing a visible marker of common appearance to the coalition.

(*iv*) As predicted, when coalition membership became marked by cues of shared appearance, the degree to which subjects encoded it increased substantially [$n = 52$, $M = 4.62$ (SD = 3.62), $t^*(51) = 9.20$, $P = 1.02 \times 10^{-17}$]. The size of the coalition effect in experiment 2 was more than 2.5 times larger than in experiment 1, which used only allegiance cues (0.79 versus 0.31). The increase in the extent to which subjects encoded coalitional alliance was both large and significant [experiment 1 versus experiment 2: $t(95.4) = -5.945$, $r = 0.52$, $P = 2.3 \times 10^{-8}$].

This indicates that endowing coalition with the attribute of visibility by an arbitrary cue was an effective manipulation for increasing categorization of targets by coalition. More importantly, it confirms prediction *3*: it shows that a new and arbitrary coalition can be encoded just as strongly as race is. Indeed, this understates the result: the arbitrary coalition in experiment 2 was encoded even more strongly than race was at its strongest [0.79 vs. 0.67) in experiment 1 ($t(91) = 3.81$, $r = 0.37$, $P = 0.00013$).

This observation allows us to address two key questions: when race ceases to be a predictor of coalitional allegiance within a given social context, and when coalition membership is marked by cues of shared appearance (as race is), does coalition acquire the robust properties race had, and does race lose the strength it once had? These predictions (nos. 1 and 4) were confirmed.

(*v*) In experiment 2, when cues of coalitional affiliation were amplified through shared appearance, subjects did continue to categorize on the basis of race [$M = 1.40$ (SD = 2.51), $t^*(51) = 4.038$, $P = 0.000091$], but the size of the race effect was diminished, from 0.67 in experiment 1, where coalition cues were subtle, to 0.49 in experiment 2, where they were amplified. This reduction in the encoding of race was significant and substantial [experiment 1 versus experiment 2: $t(105) = 1.83$, $r = 0.18$, $P = 0.035$].

(*vi*) When an alternative, and contextually relevant, coalitional categorization was reinforced through shared appearance in experiment 2, coalition was encoded far more strongly than race:

$r = 0.79$ for coalition, 0.49 for race [$t^*(51) = 6.14$, $r = 0.65$, $P = 6.3 \times 10^{-8}$]. The large effect size (0.65) associated with this difference shows that when race is irrelevant to the coalitional conflict at hand, the encoding machinery tracks coalition membership much more carefully than race.

This result demonstrates that social context—in this case, one in which an important coalitional dimension does not track or correspond to race—can diminish the extent to which race is encoded. So contrary to prior claims, race is not inevitably encoded with equal strength across social contexts. In a social world where the active coalitions are easy to encode and do not track race—even briefly—encoding by race decreases.

These findings indicate that subjects bring the tendency to categorize by race with them into the experiment, but then begin to lose it as the circuitry detects that it no longer predicts relevant coalitions in this context. In both experiments 1 and 2, race was uncorrelated with coalition, but because coalition is far easier to perceive in experiment 2 during both encoding and retrieval, the irrelevance of race to coalition is also far easier to perceive in that experiment.

It bears underlining that the racial appearance of the targets remained exactly the same in both experiments, so that if the tendency to confuse two individuals of the same race was based on their being perceptually similar, rather than being categorized abstractly as belonging to the same race, then amplifying competing coalitional cues should have had no effect. This means that racial encoding is cued by similarity of appearance, but does not consist of similarity of appearance. This supports the view that appearance cues—including race—are picked up or discarded to the extent that they predict coalition, and not because they are intrinsically salient.

(*vii*) Lastly, as cues to coalition membership were amplified (experiment 1 versus experiment 2), the *relative* importance of race versus coalition reversed itself. When coalition cues were not based on appearance, race remained more important than the newly manufactured coalition; however, when they were amplified by means of an arbitrary shared appearance cue, coalition became more important than race.

Fig. 1 depicts the extent to which these ratios deviate from the 1.0 ratio that equal weighting would produce.

There is nothing mathematically necessary about this flip. Even if amplifying coalition cues increases the extent to which subjects index a target's coalitional affiliation, it need not do so to the point where coalition becomes more important than race.

The results of two control experiments demonstrate this conceptual point. Experiments 3 and 4 were identical to experiments 1 and 2, with one exception: Instead of varying the race of targets, we varied their sex. Unlike race, sex is a good candidate for a primary dimension of person representation that our minds evolved to encode across most if not all situations.

Subjects categorized by coalition in Experiments 3 and 4, replicating the coalition results from Experiments 1 and 2. When cues to coalition had to be inferred from utterances alone, the effect size for coalition was 0.35 (comparable to 0.31 for the analogous experiment 1; [experiment 3: $n = 55$, $M = 0.98$ (SD = 2.69), $t^*(54) = 2.709$, $P = 0.0045$]; when cues were amplified, the effect size for coalition was 0.81 [comparable to 0.79 in experiment 2; experiment 4: $n = 57$, $M = 4.29$ (SD = 3.16), $t^*(56) = 10.241$, $P = 9.6 \times 10^{-15}$]. In contrast to the race results, the extent to which subjects categorized by the targets' sex was very high in both experiments: effect sizes were 0.91 and 0.84 for experiments 3 and 4, respectively [experiment 3: $M = 4.75$ (SD = 2.23), $t^*(54) = 15.81$, $P = 3.2 \times 10^{-22}$; experiment 4: $M = 3.87$ (SD = 2.53), $t^*(56) = 11.52$, $P = 1.1 \times 10^{-16}$]. Although there was a diminution in the extent to which sex was encoded when coalitional cues were amplified [from 0.91 to 0.84; $t(110)=1.94$, $r = 0.18$, $P = 0.027$], the extent to which subjects categorized by sex remained very large (0.84). In sharp contrast to the relation between coalition and race in experiment 2, sex was always encoded more strongly than coalition (Fig. 1).

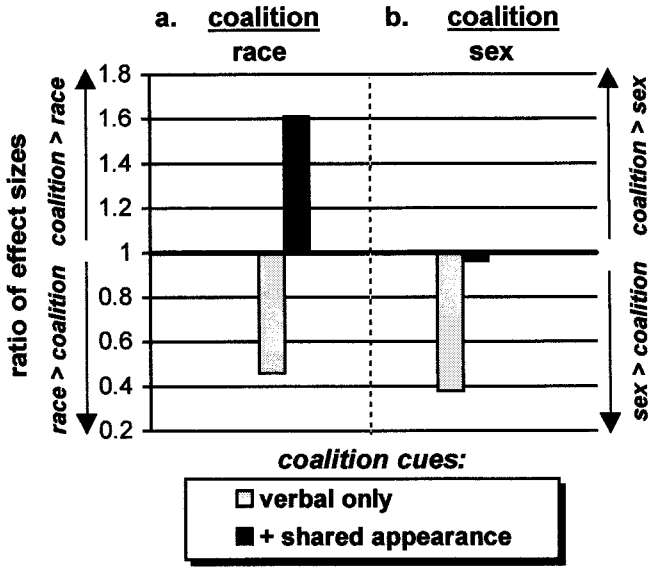

Fig. 1. (*a*) Relative importance of coalition versus race in social categorization. Effect sizes index how strongly subjects categorized along a dimension. If coalition and race were given equal weight, then the ratio of their effect sizes (coalition/race) would be ≈1.0. It would be <1.0 if race were weighted more heavily than coalition, and >1.0 if coalition were weighted more heavily than race. In experiment 1, where the only cues to coalition membership were verbal, this ratio was 0.46, indicating that subjects were more strongly encoding race (coalition = 0.31, race = 0.67). But in experiment 2, where coalitional cues were amplified by the addition of shared appearance cues, the coalition to race ratio was 1.61 (coalition = 0.79, race = 0.49), indicating that coalitional alliance was encoded more strongly than race. The same is not true for coalition versus sex. (*b*) Relative importance of coalition versus sex. When coalition cues were amplified, subjects continued to strongly categorize on the basis of sex, and the targets' sex and coalition membership were weighted about equally.

Moreover, the effect sizes for sex were significantly larger than those for race in the analogous conditions (confirming prediction 5): 0.91 for sex versus 0.67 for race when only verbal allegiance cues were present (experiment 3 vs. experiment 1: $Z = 3.66$, $r = 0.35$, $P = 0.00013$), and 0.84 for sex versus 0.49 for race when coalitional cues were amplified by appearance [experiment 4 vs. experiment 2: $Z = 3.53$, $r = 0.33$, $P = 0.00021$; the same difference is supported by t tests comparing sex and race: experiment 3 vs. 1: $t(106) = 5.37$, $r = 0.46$, $P = 2.4 \times 10^{-7}$; experiment 4 vs. 2: $t(106) = 5.10$, $r = 0.44$, $P = 7.5 \times 10^{-5}$; see Fig. 2. The hypothesis that sex is a true primary categorical dimension—encoded in an automatic and (relatively) mandatory fashion—is supported by these results: the effect sizes for sex are very large, and they stay large even when coalitional cues are amplified.

Although there is nothing about the protocol that prevents subjects from encoding a number of different dimensions at once, one might reasonably hypothesize that there exists a domain-general attentional constraint creating an inevitable tradeoff between the degree to which different dimensions are encoded. On this view, the inverse relationship between race and coalition does not imply any special relationship between race and coalition: any increase in one channel is expected to be compensated for by decreased encoding in another. These data falsify such a view. Tellingly, sex was encoded far more strongly than race (in experiments 3 and 4 versus 1 and 2 and 5 and 6), yet this did not lead to any reduced encoding of coalition, as a tradeoff view would predict. In fact, the strength of coalition

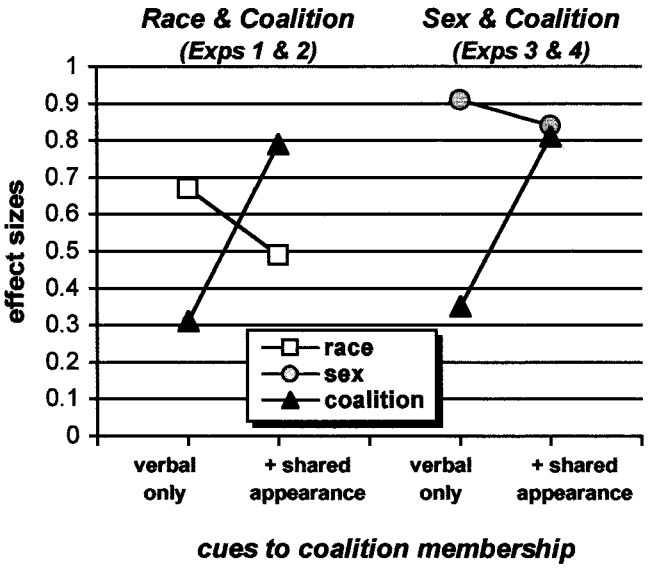

Fig. 2. Strength of categorization by coalition, race, and sex. Coalition was encoded even in the absence of shared appearance cues. Sex was always encoded more strongly than race. When coalition membership was amplified by the addition of shared appearance cues (which are always present for race and sex), then coalition was encoded as strongly as sex, and far more strongly than race. The data suggest that coalition and sex are primary dimensions of person representation, whereas race is not.

encoding increased slightly in both cases, from 0.31 to 0.35 [verbal only: experiment 1 (race) versus experiment 3 (sex)], and from 0.79 to 0.81 [amplified coalition cues: experiment 2 (race) versus experiment 4 (sex)]. This means there was attentional capacity to spare in the race experiments, eliminating attentional constraint as a credible explanation for the diminution in race encoding observed in experiment 2.

The results for sex differ in another important way from those for race. In the race experiments, the relative importance of coalition and race sharply reversed when coalition cues were amplified. This did not happen in the sex experiments. Although the importance of coalition increased in experiment 4, approaching the high levels of encoding accorded to sex, it did not outstrip sex (coalition to sex ratio: 0.81 to 0.84 = 0.96). (See Fig. 1 and *Appendix*.) In other words, amplifying coalition cues does not cause coalition to be encoded more strongly than sex. But amplifying coalition cues does cause coalition to be encoded more strongly than race: the relative importance of race decreased dramatically as cues to coalition membership became more obvious—indeed, the coalition effect was >60% larger than the race effect. This is what one would expect if race were a proxy for coalition, but sex were not. This can be seen even more strikingly in experiments 5 and 6.

To check the robustness of our results, we conducted experiments 5 and 6 as exact replications of experiments 1 and 2, but with photographs of new individuals. Although there were minor differences in the results, the same basic pattern for race obtained with the new stimuli: when cues to coalitional alliance were verbal only, the size of the race effect was large (0.57) and significant [$n = 51$, $M = 1.86$ (SD = 2.69), $t^*(50) = 4.94$, $P = 9.1 \times 10^{-6}$]. But when cues to coalitional alliance were augmented visually, the size of the race effect dropped substantially, to 0.15. This drop in effect sizes—from 0.57 to 0.15—is itself significant ($Z = 2.45$, $r = 0.24$, $P = 0.0073$), as is a direct test of the race effect in these two conditions [$t(101) = 2.95$, $r = 0.28$, $P = 0.004$]. Indeed, in the condition in which coalitional cues were amplified, there was no statistically significant tendency for subjects to encode the race of targets [$n = 52$, $M = 0.37$ (SD = 2.45), $t^*(51) = 1.07$, $P = 0.29$]. In this condition, it would appear that the extent to which subjects encoded targets by their race was not merely diminished, it was erased.

Conclusions

What is most striking about these results is just how easy it was to diminish the importance of race by manipulating coalition—especially given the repeated failure over decades to find other means to influence racial encoding. The sensitivity of race to coalitional manipulation lends credence to the hypothesis that, to the human mind, race is simply one historically contingent subtype of coalition. Our subjects had experienced a lifetime in which ethnicity (including race) was an ecologically valid predictor of people's social alliances and coalitional affiliations. Yet less than 4 min of exposure to an alternative social world in which race was irrelevant to the prevailing system of alliance caused a dramatic decrease in the extent to which they categorized others by race. This implies that coalition, and hence race, is a volatile, dynamically updated cognitive variable, easily overwritten by new circumstances. If the same processes govern categorization outside the laboratory, then the prospects for reducing or even eliminating the widespread tendency to categorize persons by race may be very good indeed.

Appendix

Subjects. All subjects were undergraduates at the University of California, Santa Barbara, average age = 19. In each experiment, the sex ratio of subjects was ≈50:50. Each experiment tested individuals who had not participated in any of the other experiments. Subjects were primarily Euro- (including Hispanic) and Asian-American.

Photographic Stimuli. Each photograph used was a front-facing color photo of the head and upper body of a young man or woman wearing a plain basketball jersey. In experiments 1, 3, and 5, all individuals were wearing jerseys of the same color, with no identifying marks. In experiments 2, 4, and 6, each coalition wore jerseys of a different color (colors were manipulated by using Adobe PHOTOSHOP; Adobe Systems, Mountain View, CA). Thus the target photos were held constant across experiments 1–4. Experiments 5 and 6 used photos of a different set of individuals.

Memory Confusion Protocol. Subjects were told that (*i*) they will be seeing a series of photographs of individuals, each of which is paired with a sentence uttered by the individual pictured (sequentially displayed on a computer terminal), (*ii*) each pictured individual belongs to one of two rival basketball teams that had been in a fight during the previous season, and their sentences were uttered in the context of a group conversation, and (*iii*) their task is to form an impression of the target individuals as they are viewed. Each subject viewed 24 sentences for 8.5 s each; each sentence was paired with a photo of one man (thus each man uttered a total of three sentences). The pairing of individual photos with sentences was randomized across subjects. The sentences, whose content was antagonistic and coalitional, were presented as if they were sequential statements in a heated conversation. After viewing all of the photos and sentences, subjects performed a 1-min distracter task. After completing that task, an array composed of all of the photos of the previously displayed target individuals appeared on the screen. One at a time, each sentence that the subject had viewed previously appeared on the screen (in random order). The subjects' task at this point was to recall which target individual said each particular sentence. This is a difficult task, and subjects made many errors (this is expected and necessary, as the method depends on an analysis of errors). If they are categorizing the target individuals into different groups on the basis of some cue, then this

should be reflected in the type of errors they make. [Errors in this task could, in principle, arise from a failure to encode the properties of the speaker of the sentence at viewing time or from a failure to use encoded information at the time of recall. Although this method cannot distinguish between these possibilities, it seems more likely that errors are caused by failures of encoding. It is not obvious why encoded information would not be used at the time of recall if it were available. (Instructions were identical across experiments.) Regardless of where the information is lost in this chain, the result is that coalitional situations that are not correlated with race diminish the use of racial information.]

In experiments 1, 2, 5, and 6, the teams were each composed of two Euro-American players and two African-American players. For each subject, the computer program constructed two coalitions (teams) randomly, with the provision that the players be equally divided racially. Teams alternated sentences and players were assigned randomly to sentences, with the constraint that the first four sentences consisted of first two Euro-American speakers, and then two African-American speakers (or vice versa). This restriction was added to reinforce the multiracial aspect of the two coalitions from the beginning of the dialogue. Experiments 3 and 4 were designed in the same way, except that Euro-American women were used instead of African-American men.

The antagonistic content of the sentences, and the order in which they were uttered, contain sufficient information to infer which team each individual is on. To independently check this variable, we gave the sentences alone (on paper), each paired with a name, to a different set of subjects, with explicit instructions to figure out which person was on which team (subjects in experiments 1–6 were never instructed to attend to or infer team membership). The majority (70%) of subjects scored 100% correct, and all but one of the remaining subjects made only one error.

Statistical Analyses. Because there were eight photos evenly divided along two dimensions, it is possible to make four different types of error. Subjects, in misattributing a sentence to a player, could pick another player on the same team and of the same race (one photo), a player on the same team but of a different race (two photos), a player on the other team of the same race (two photos), or a photo of a player on the other team of a different race (two photos). To compensate for the lower prior probability of making a same-team/same-race error, the error rates of the other three categories were divided in half before any of the following statistical tests were conducted. This correction for different base rates is necessary, and standard in the literature (13–15).

To test for coalition effects, the number of (same coalition, same race + same coalition, different race) errors was compared with the number of (different coalition, same race + different coalition, different race) errors. To test for race effects, the number of (same race, same coalition + same race, different coalition) errors was compared with the number of (different race, same coalition + different race, different coalition) errors. For experiments 3 and 4, one makes the equivalent computations, but substituting sex for race.

In all of the tests to assess whether a particular dimension had been encoded, the data point for each subject is the number of within category errors minus the number of between category errors made by that individual, and the hypothesis being tested is that this is greater than zero. Thus these are t^* tests (paired t tests). Because we had prior predictions, significance tests were one-tailed, unless otherwise noted.

Using the results of a t^* test, one can index how strongly subjects were encoding a dimension by computing an effect size, r, which varies from 0 to 1 (28). The more within-category errors (compared with between-category errors) that subjects make, the larger the effect size, r, will be. By comparing effect sizes, one can see whether subjects encoded a dimension more strongly in one condition than in another, even if they encoded it to a significant extent in both.

Sex Differences. Because this article is about the encoding of race, and because none of our conclusions about race change when the data are analyzed separately for male and female subjects, we have collapsed over the two sexes for the purposes of this article. However, an analysis of selection pressures in the context of coalitions led us to predict that machinery for detecting multi-individual coalitions and alliances—especially competitive ones—will be present in both sexes, but easier to activate in men than in women (25). There are data that support these predictions (R.K., L.C., and J.T., unpublished data).

Note 1. The t^* tests do not compare aggregated means (they are within-subject tests); this is to ensure that results are general, i.e., not caused by a minority of subjects with extreme scores. An $M = 0.90$ for coalition represents ≈20% more within than between category errors (aggregated Ms: 4.61 between; 5.51 within); in experiment 2, $M = 4.62$ for coalition represents ≈200% more within than between category errors (aggregated Ms: 2.31 between; 6.92 within). (Range of aggregated means across experiments: within-category errors: 5.32–7.25; between-category errors: 2.31–4.61.)

Funding for this project was provided by the Harry Frank Guggenheim Foundation, the James S. McDonnell Foundation, National Science Foundation Grant BNS9157-449 (to J.T.), and the University of California, Santa Barbara Office of Research and Academic Senate.

1. Brewer, M. (1979) in *The Social Psychology of Intergroup Relations*, eds. Austin, W. G. & Worchel, S. (Brooks/Cole, Monterey, CA), pp. 71–84.
2. Brewer, M. (1979) *Psychol. Bull.* **86,** 307–324.
3. Sherif, M., Harvey, O., White, B., Hood, W. & Sherif, C. (1961) *Intergroup Conflict and Cooperation* (Univ. of Oklahoma Book Exchange, Norman).
4. Tajfel, H., Billig, M., Bundy, R. & Flament, C. (1971) *Eur. J. Soc. Psychol.* **1,** 149–178.
5. Tajfel, H. (1978) *Differentiation Between Social Groups* (Academic, New York).
6. Billig, M. & Tajfel, H. (1973) *Eur. J. Soc. Psychol.* **3,** 27–52.
7. Locksley, A., Ortiz, V. & Hepburn, C. (1980) *J. Pers. Soc. Psychol.* **39,** 773–783.
8. Levine, R. & Campbell, D. (1972) *Ethnocentrism* (Wiley, New York).
9. Messick, D. & Mackie, D. (1989) *Annu. Rev. Psychol.* **40,** 45–81.
10. Hamilton, D., Stroessner, S. & Driscoll, D. (1994) in *Social Cognition: Impact on Social Psychology*, eds. Devine, P., Hamilton, D. & Ostrom T. (Academic, San Diego), pp. 291–321.
11. Brewer, M. (1988) *Adv. Soc. Cognit.* **1,** 1–36.
12. Fiske, S. & Neuberg, S. (1990) *Adv. Exp. Soc. Psychol.* **23,** 1–74.
13. Taylor, S., Fiske, S., Etcoff, N. & Ruderman, A. (1978) *J. Pers. Soc. Psychol.* **36,** 778–793.
14. Hewstone, M., Hantzi, A. & Johnston, L. (1991) *Eur. J. Soc. Psychol.* **21,** 517–528.
15. Stangor, C., Lynch, L., Duan, C. & Glass, B. (1992) *J. Pers. Soc. Psychol.* **62,** 207–218.
16. Kelly, R. (1995) *The Foraging Spectrum* (Smithsonian Inst., Washington, DC).
17. Lewontin, R. (1972) *Evol. Biol.* **6,** 381–398.
18. Nei, M. & Roychoudhury, A. (1993) *Mol. Biol. Evol.* **10,** 927–943.
19. Keeley, L. (1996) *War Before Civilization* (Oxford Univ. Press, Oxford).
20. Ember, C. R. (1978) *Ethnology* **27,** 239–448.
21. Manson, J. & Wrangham, R. (1991) *Curr. Anthropol.* **32,** 369–390.
22. Chagnon, N. (1992) *Yanomamo* (Harcourt, Fort Worth, TX), 4th Ed.
23. Smuts, B., Cheney, D., Seyfarth, R., Wrangham, R. & Struhsaker, T. (1987) *Primate Societies* (Univ. of Chicago Press, Chicago).
24. Kurzban, R. (1998) Ph.D. dissertation (Univ. of California, Santa Barbara).
25. Tooby, J. & Cosmides, L., eds. (2002) *Evolutionary Psychology: Foundational Papers* (MIT Press, Cambridge, MA), in press.
26. Sidanius, J. & Pratto, F. (1999) *Social Dominance* (Cambridge Univ. Press, New York).
27. Hirschfeld, L. (1996) *Race in the Making* (MIT Press, Cambridge, MA).
28. Rosenthal, R. & Rosnow, R. (1991) *Essentials of Behavioral Research* (McGraw–Hill, Boston), 2nd Ed.