The Coevolution of Cultural Groups and Ingroup Favoritism — **Source link**

Charles Efferson, Charles Efferson, Rafael Lalive, Ernst Fehr

Institutions: University of Zurich, Santa Fe Institute, University of Lausanne

Topics: In-group favoritism, Ingroups and outgroups, Cultural group selection, Population and Social group

Related papers:

- Social categorization and intergroup behaviour

- Shared norms and the evolution of ethnic markers

- Parochial altruism in humans

- The Coevolution of Parochial Altruism and War

- Five Rules for the Evolution of Cooperation

Share this paper:

The Coevolution of Cultural Groups and Ingroup Favoritism

Charles Efferson, *et al.*

Science **321**, 1844 (2008);
DOI: 10.1126/science.1155805

were inconspicuous on routine histological inspection (Fig. 3B, top left). Staining with antibody to K8 facilitated the detection of the ectopic foci and confirmed their epithelial origin, whereas in the lungs of uninjected mice only the bronchial epithelium was stained (Fig. 3B, right).

To determine whether the ectopic foci of normal epithelial cells persist and grow in the foreign environment of the lung, we counted the total number of discrete foci in lung sections at different times after injection and looked at proliferation markers in these foci. The total number of foci found in lung sections from C57BL6/J recipients injected with 4×10^5 syngeneic mammary cells was similar in the animals surveyed at 3 weeks ($n = 3$ mice) and those surveyed at 10 weeks ($n = 3$ mice) after injection (42 ± 7 and 56 ± 22 in 10 paraffin lung sections, respectively). Moreover, the efficiency with which the wild-type cells were able to form these small epithelial clusters was similar to the efficiency with which we were able to induce ectopic tumors after injecting cells from doxycycline-naïve *TOM;TOR;MTB* donors [1.2 ± 0.4 (SD) versus 1.7 ± 1.4 (SD) per 10,000 cells injected, $n = 6$ and 8 mice, respectively; measured as described in (*14*)]. This result strongly argues that most or all of the mammary cells that are capable of surviving in the lung are able to respond to the initiating oncogene expression by forming an ectopic mammary tumor.

In both nontransgenic C57BL6/J– and β-*actin*-GFP–derived foci, occasional cells displayed mitotic activity (Fig. 3C). Consistent with this result, the green foci found under excitation light in the lungs of animals injected with mammary cells from β-*actin*-GFP mice 16 weeks after injection were larger in size than those found in recipients of the same preparation 1 week after injection (Fig. 3D). Ectopic epithelial outgrowths contained K8- and SMA-positive cells, such as observed in intact mammary glands (fig. S4A), and the outgrowths occasionally displayed a glandular appearance. Despite prolonged residence in the lung (up to 4 months), the green cells recovered from the recipients' lungs were competent to form hollow acinar structures in three-dimensional morphogenesis assays (fig. S4B) and secondary mammary outgrowths in cleared fat pads of *Rag1*$^{-/-}$ females (Fig. 3E). These findings establish that the ectopic cells residing in the lungs are indeed of mammary origin, that they are viable and mitotically active, and that at least some of them are multipotent and able to support full mammary development.

The experiments described here show that, in the absence of an active oncogene, dissociated cells from an untransformed mouse mammary gland can establish residence in the ectopic environment of the lung, grow slowly, and remain clinically undetectable after IV injection. The same cells can give rise to metastatic malignancies upon activation of oncogenes that can produce mammary tumors in an intact gland. It is widely acknowledged that multiple steps are required to establish metastases, including intravasion of cells from primary tumors into blood vessels or lym-

phatics; survival in the circulation, extravasation, and establishment of cells at ectopic sites; and malignant growth. Because we have injected mammary cells from transgenic mouse donors into tail veins of recipient mice, we have not examined the requirements for intravasation. We have, however, demonstrated that activated oncogenes and cellular transformation are not required for any of the subsequent steps, save for malignant growth at ectopic sites. These findings indicate that properties inherent in normal cells are sufficient for negotiating a substantial portion of the metastatic cascade. Considerable experimental and clinical evidence favors the idea that cells from small cancers may spread to distant sites early in tumorigenesis and account for dormancy and late relapse in human breast cancer (*2*, *18*). Although we do not know whether premalignant cells can enter the systemic circulation during these early stages and become sources of later metastatic tumors, our observations argue that this hypothesis should be tested. The finding that metastatic disease can arise from untransformed mammary cells in the circulation refines our conception of cancer progression, and suggests that each step in the metastatic cascade should be examined to establish its functional requirements, including those performed by normal cells. Such functions might be susceptible to inhibitory strategies that can ablate disseminated pre-malignant or malignant cells and thereby diminish the mortality caused by cancer.

References and Notes
1. D. Hanahan, R. A. Weinberg, *Cell* **100**, 57 (2000).
2. Y. Husemann *et al.*, *Cancer Cell* **13**, 58 (2008).
3. G. P. Gupta *et al.*, *Nature* **446**, 765 (2007).
4. P. B. Gupta *et al.*, *Nat. Genet.* **37**, 1047 (2005).
5. K. A. Hartwell *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18969 (2006).
6. M. Jechlinger *et al.*, *J. Clin. Invest.* **116**, 1561 (2006).
7. R. S. Muraoka *et al.*, *J. Clin. Invest.* **109**, 1551 (2002).
8. Y. Kang *et al.*, *Cancer Cell* **3**, 537 (2003).
9. A. Muller *et al.*, *Nature* **410**, 50 (2001).
10. J. Yang *et al.*, *Cell* **117**, 927 (2004).
11. A. E. Karnoub *et al.*, *Nature* **449**, 557 (2007).
12. R. Bernards, R. A. Weinberg, *Nature* **418**, 823 (2002).
13. K. Podsypanina, K. Politi, L. J. Beverly, H. E. Varmus, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5242 (2008).
14. Materials and methods are available as supporting material on *Science* Online.
15. P. Mombaerts *et al.*, *Cell* **68**, 869 (1992).
16. E. J. Gunther *et al.*, *FASEB J.* **16**, 283 (2002).
17. M. Okabe, M. Ikawa, K. Kominami, T. Nakanishi, Y. Nishimune, *FEBS Lett.* **407**, 313 (1997).
18. J. A. Aguirre-Ghiso, *Nat. Rev. Cancer* **7**, 834 (2007).
19. We thank M. A. Melnick, G. Sanchez, A. Giannakou, and J. Demers for expert handling of the mouse colony; L. Chodosh for providing *MMTV-rtTA* transgenic mice; D. Felsher and J. M. Bishop for providing *TetO-MYC* transgenic mice; A. Olshen for assistance with statistical analysis; and L. K. Tan for assistance with histological analysis. Supported in part by awards from NIH (K01 CA118731 to K.P., P01 CA94060 to H.V., and R24 CA83084 and P30-CA 08748, which provides partial support for core facilities used in conducting this investigation), the Martell Foundation (to H.V.), and the U.S. Department of Defense (W81XWH-05-1-0220 to M.J.).

The Coevolution of Cultural Groups and Ingroup Favoritism

Charles Efferson,[1,2]* Rafael Lalive,[3] Ernst Fehr[1,4]

Cultural boundaries have often been the basis for discrimination, nationalism, religious wars, and genocide. Little is known, however, about how cultural groups form or the evolutionary forces behind group affiliation and ingroup favoritism. Hence, we examine these forces experimentally and show that arbitrary symbolic markers, though initially meaningless, evolve to play a key role in cultural group formation and ingroup favoritism because they enable a population of heterogeneous individuals to solve important coordination problems. This process requires that individuals differ in some critical but unobservable way and that their markers be freely and flexibly chosen. If these conditions are met, markers become accurate predictors of behavior. The resulting social environment includes strong incentives to bias interactions toward others with the same marker, and subjects accordingly show strong ingroup favoritism. When markers do not acquire meaning as accurate predictors of behavior, players show a markedly reduced taste for ingroup favoritism. Our results support the prominent evolutionary hypothesis that cultural processes can reshape the selective pressures facing individuals and so favor the evolution of behavioral traits not previously advantaged.

A cultural group is a group of people who share a set of beliefs, behavioral norms, and behavioral expectations that is recognizably different from those of other groups (*1*). Beliefs, norms, and expectations, however, are often not directly observable, and so by themselves they do not provide a practical basis for identifying cultural groups in everyday social interactions. Nonetheless, cultural groups are frequently identifiable through ethnic markers, which are arbitrary but observable traits like dress, dialect, and body modification that symbolically and conspicuously signal group affiliation (*1*–*5*).

Symbolic traits of this sort can be crucial to social and economic outcomes. When ethnic markers covary with other cultural traits, individuals can

potentially use markers to everyone's mutual advantage as indicators of what would otherwise be unobservable variation in beliefs, norms, and expectations. More nefariously, ethnic markers can lead to segregation, ethnic discrimination, and persistent inequality, even in the paradoxical cases when everyone prefers integration (6–8) or when ethnicity indicates nothing about competence in a given domain (9, 10). Indeed, parochialism and prejudice often mar intergroup relations. People show favoritism toward ingroup members and indifference, hostility, or mistrust toward outgroup members (11–19). They do so even when groups are transient and group boundaries rest on the flimsiest of distinctions among individuals (15, 20–22). These findings have potentially broad significance because recent theoretical research has closely and surprisingly tied outgroup hostility to the evolution of human prosociality within groups (23, 24).

None of this, however, explains how a group gets to be a group and why. The long tradition of empirical research on intergroup relations (11–22, 25) includes two basic approaches to defining groups. Studies have either used preexisting cultural groups, which formed beyond the ken of the studies in question, or subjects were assigned to groups exogenously as part of an experiment involving the effects of social categorization. These methods can be powerful for many questions (12, 16), but they cannot expose the mechanisms behind the formation of cultural groups. These mechanisms also represent a gap in evolutionary theories of human prosociality. Although the initial evolution of cultural groups may have little to do with cooperation, much of the theory on the evolution of human prosociality relies heavily on the observation that human populations are subdivided into cultural groups (23, 24, 26). This theoretical work, however, simply imposes the required population structure exogenously. The endogenous formation of cultural groups represents a plausible route to the required population structure that figures prominently but remains unexplained in evolutionary theories of human prosociality.

We conducted a set of experiments to identify the conditions required for cultural groups to form endogenously and for subjects to show ingroup favoritism in their subsequent social interactions. We used neither preexisting cultural groups nor groups created exogenously by the experimenter. Our task instead was to see if and when symbolically marked groups form endogenously and whether their formation can lead to a preference for interactions with others having the same symbolic marker. This preference was our operational measure of ingroup favoritism in the experiment, and more generally such preferences

can limit social interactions across cultural boundaries and potentially play a key role in the development of ethnocentric attitudes (27). If such a preference were to emerge endogenously in our setting, the result would support a central hypothesis in evolutionary social science (27–31). This hypothesis posits that a cultural evolutionary process can modify the selective environment facing individuals and so lead to the evolution, whether cultural or genetic, of traits that were not previously advantageous. In our case, the question is whether the evolution of cultural groups during an experiment can reconstitute the social environment to benefit ingroup favoritism in a way that did not obtain at the beginning of the experiment.

In theory, cultural groups form when variation in an unobservable but socially critical variable becomes manifest. Consider a population of players playing a simultaneous, two-person coordination game with multiple equilibria. Players can choose behavior A or B. If two players meet and choose the same behavior, a large payoff results. If they choose different behaviors (32), a small payoff results. Some players expect to coordinate on A, others on B. If players with different expectations meet, an information problem results. One simply has to play the odds and risk miscoordinating with someone who has incompatible expectations. This kind of problem is general. Variation in behavioral norms and expectations is widespread (1, 33, 34), and the mixing of people with different expectations occurs frequently (1, 35, 36). This mixing, however, creates the potential for people with discordant social expectations to meet, interact, and miscoordinate. Variation in expectations, however, is not enough for the existence of cultural groups because this variation is not directly observable.

Symbolic markers can change matters greatly, but only if they covary with expectations and by extension behavior. To illustrate, let players in our coordination game wear shirts with either
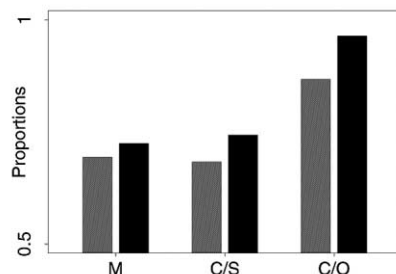
triangles or circles. The shape on one's shirt does not affect payoffs, and so it fills the theoretical role of a symbolic marker. Consider a hypothetical population of 100 people, 50 of whom expect to coordinate on A and 50 on B. In addition, the 50 players who expect to coordinate on A have triangles on their shirts, and the 50 players who expect to coordinate on B have circles on their shirts. The distribution of behavior-marker types in the population is consequently 50 (A, ▲) individuals, 0 (A, ●) individuals, 0 (B, ▲) individuals, and 50 (B, ●) individuals. The covariation between behavior and marker is at its maximum possible value in this example, and the markers perfectly reveal expectations and their associated behaviors in the coordination game. More generally, when covariation characterizes the distribution of behavior-marker types, the observable markers allow one to draw statistical inferences about what is unobservable but really important, namely, behavioral expectations in a social setting with multiple equilibria. When this is true, interacting preferentially with others having the same marker reduces the probability of miscoordination and
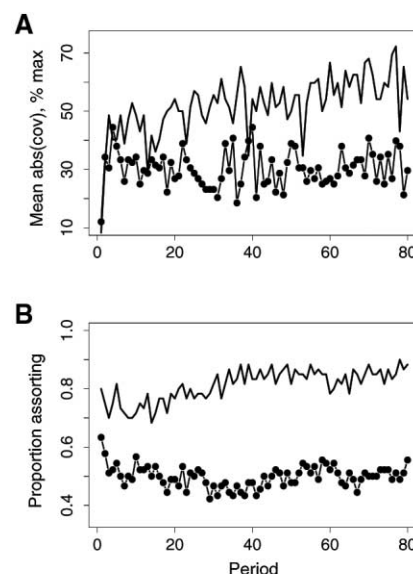
Fig. 2. (**A**) The informational content of the marker. The graph shows the mean magnitude of the covariance between behavior and marker in a subpopulation relative to the theoretical maximum for the marker-randomized (line with filled circles) and the marker-maintained (solid line) treatments. The period trend for marker-randomized is not significant [Newey-West (40) regression, maximal lag of 10, t test, P = 0.368], whereas it is highly significant for the marker-maintained treatment (Newey-West, lag of 10, t test, P < 0.001). (**B**) Ingroup favoritism, as indicated by the proportion of players requesting a partner with the same shape. The marker-randomized period trend is not significant (Newey-West, lag of 10, t test, P = 0.868). The marker-maintained period trend is highly significant (Newey-West, lag of 10, t test, P < 0.001), leading to large differences in ingroup favoritism across treatments.

Fig. 1. Summary of linked choices for the marker-randomized (gray) and marker-maintained (black) treatments. The behavior and marker chosen in stage 1 are coded as either linked or unlinked relative to the behavior and marker chosen in stage 1 of the previous period. Proportions are plotted for the cases in which the player miscoordinated (M) in the previous period, coordinated on the suboptimal (C/S) behavior (i.e., A in subpopulation 2 or B in subpopulation 1), and coordinated on the optimal (C/O) behavior (i.e., A in subpopulation 1 or B in subpopulation 2).

[1]Institute for Empirical Research in Economics, University of Zürich, Blümlisalpstrasse 10, 8006 Zürich, Switzerland. [2]Santa Fe Institute, NM 87501, USA. [3]Department of Economics, University of Lausanne, 1015 Lausanne, Switzerland. [4]Collegium Helveticum, 8092 Zürich, Switzerland.

*To whom correspondence should be addressed. E-mail: efferson@iew.uzh.ch

increases expected payoffs. The puzzle, however, is how to get strong covariation endogenously in decentralized societies under limited information about the distribution of behavior-marker combinations. How does symbolic meaning emerge in the absence of fiat? Interestingly, mixing players with different expectations, which creates the original problem, also creates a potential solution. It does so by producing small amounts of covariation (37) that can feed back into the system and accumulate dynamically (38, 39).

The accumulation of covariation requires more than mixing, however, because mixing by itself often creates only a small amount of covariation between behavior and marker (38). During our experiment, individuals did not have information about the aggregate distribution of behavior-marker combinations, and thus it would have been difficult or impossible to recognize an initially weak relation between behavior and marker. Covariation can increase, however, if individuals link behaviors and markers in specific ways. Linkage refers to a tendency for an individual either to retain both her current behavior and marker or to change both her behavior and marker; what an individual does not do is change one trait but not the other. Linkage is crucial because it preserves the covariation created by earlier mixing, while continued mixing creates additional covariation that feeds back into the system and gets added to existing covariation. The result is that the total covariation accumulates, and this increases the economic incentives to interact with others having the same marker. For covariation to accumulate, however, linkage should not be indiscriminate. Rather, theory suggests it should be more prevalent in specific situations like those in which individuals acquire information about economically successful behavior-marker combinations (38, 39). If individuals, however, never link under any circumstances because they choose behaviors and markers independently, covariation is constantly destroyed, and markers cannot become strongly associated with behavior.

We conducted the following experiment to see if players would show a preference for (i) linking behaviors and markers and for (ii) interacting with partners displaying the same marker. In addition, we wanted to know (iii) whether linkage, if present, would generate sizable covariation between behavior and marker, which would then enable subjects to increase coordination via ingroup favoritism. Players were assigned to one of multiple populations of 10. We randomly subdivided these 10 players into two subpopulations of 5. Players within a subpopulation played one of two coordination games (table S1). Each game had two pure-strategy equilibria, and thus players had to solve a coordination problem. Both games had two behaviors to choose from, A and B, but in subpopulation 1, coordinating on A (41 points for each player paired with another playing A) was better than coordinating on B (21 points for each player paired with another playing B), whereas in subpopulation 2, coordinating on B (41 points) was better than coordinating on A (21 points). Mis-

coordinating in either subpopulation brought a small payoff (1 point). Payoffs were designed to draw players in different subpopulations toward different behaviors and so mimic the variation in norms, preferences, and expectations that often exists because of historical separation or important but unobservable environmental differences.

To create a persistent coordination problem, players from the different subpopulations were mixed, and they were never told to which subpopulation they were assigned. If players had remained in their initial subpopulations, the game would have posed little problem. Players would have soon figured out their respective situations, and presumably players in subpopulation 1 would have only chosen A, whereas players in subpopulation 2 would have only chosen B. Each period, however, a randomly selected player from subpopulation 1 and a randomly selected player from sub-

population 2 switched subpopulations. All players knew this would happen, but no one knew which two players had switched. In sum, each player had a strong incentive to develop accurate expectations about her current subpopulation, but from time to time she found herself in a new situation where her social expectations ran askew of local norms.

Players could also condition social interactions on symbolic markers. In each period, each player chose one of two shapes, ▲ or ●. A player's payoff did not directly depend on her shape, but players could use shapes to influence with whom they would play the coordination game (39). The experiment lasted 80 periods. Each period proceeded as follows.

Stage 1. Each player chose a payoff-relevant behavior, A or B, for the coordination game and a payoff-irrelevant shape, ▲ or ●.

Stage 2. An unidentified player from each subpopulation switched subpopulations.
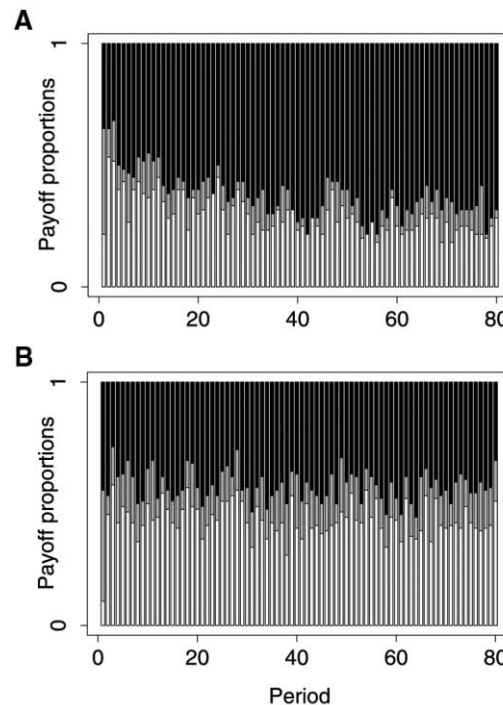
Fig. 3. Payoff proportions in the marker-maintained treatment (**A**) and the marker-randomized treatment (**B**). The graphs show the distribution of players by period coordinating on the optimal behavior (black) given the subpopulation (A in 1, B in 2), coordinating on the suboptimal behavior (gray) given the subpopulation (A in 2, B in 1), or miscoordinating (white). See supporting online text for a multinomial regression analysis.

Fig. 4. Ingroup favoritism for the modified marker-maintained (solid line), payoff-equivalent (dashed line), and fixed-marker (line with open circles) treatments. Newey-West (40) regressions indicate that the modified marker-maintained treatment began with more assortment than the other two treatments, and the differences across treatments increased through time. Comparison of regression results for the modified marker-maintained and payoff-equivalent treatments shows that the intercepts are significantly different (z test, $P < 0.001$), as are the period trends (z test, $P < 0.001$). Comparison of results for modified marker-maintained and fixed-marker shows that both the intercepts (z test, $P < 0.001$) and period trends (z test, $P = 0.019$) are significantly different.
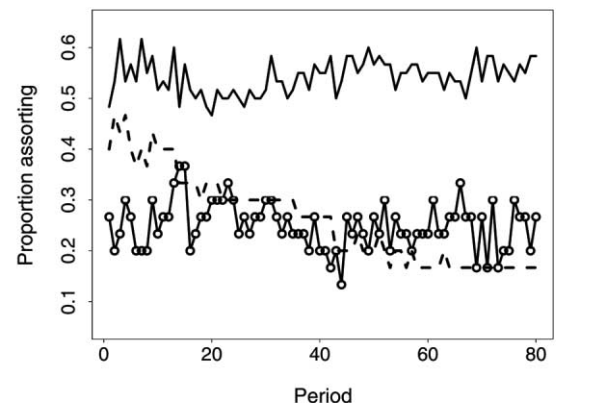
Stage 3. Each player indicated whether she wanted to play the coordination game with (i) a randomly selected player with the same shape from her subpopulation or (ii) any randomly selected player from her subpopulation.

Stage 4. Each player was paired using her choice in stage 3 and received a payoff based on her behavior, her partner's behavior, and their subpopulation.

To clarify our discussion of the results, when there was little or no covariation between behavior and marker, we will call a set of individuals who shared the same marker a "trivial" group. These groups were trivial in the sense that the markers partitioned the population into circles and triangles, but these markers did not reliably reflect any underlying variables affecting payoffs. We will call a group "cultural," in contrast, only when a set of individuals shared the same marker after a sustained increase in the aggregate covariation between behavior and marker. Groups were cultural in this case because markers did not simply partition the population into circles and triangles; they also, on average, partitioned the population into those who expected to coordinate on A versus those who expected to coordinate on B.

The experiment consisted of two treatments. In the marker-randomized treatment, each player was randomly assigned a shape after stage 2 regardless of the shape chosen in stage 1. In the marker-maintained treatment, each player retained her chosen shape. The marker-randomized treatment was a control treatment in which marker randomization precluded the possibility of the marker becoming an accurate predictor of behavior. The comparison between the two treatments shows (i) how much informational content the marker acquired in the marker-maintained treatment beyond the baseline when markers were randomly assigned and (ii) whether any differences in informational content translated into differences in the preference for ingroup favoritism.

Subpopulations were not equivalent to symbolically marked groups, whether trivial or cultural. In a given period, a player's subpopulation was the pool of players available for social interaction. A symbolically marked group, in contrast, was the set of players from the entire population with the same marker. In short, the division of players into two subpopulations, one favoring behavior A and the other behavior B, sustained variation in norms and expectations. This variation, however, was not observable, and so it could not by itself serve as a means of distinguishing one group from another. Symbolic markers, in contrast, were observable traits, and they could serve as a means of distinguishing one group from another. Markers, however, did not bear any necessary relation to behavior and subpopulation. The significance of markers, in essence, could only emerge during the experiment as a result of player choices. Markers had the potential to become the basis for determining cultural group affiliation ex post, and indeed that was our question, but they were devoid of content ex ante.

For a sustained increase in covariation, individuals have to link the behavioral and marker

dimensions. We coded behavior-marker choices from stage 1 of periods 2 to 80 as "linked" or "unlinked." A linked choice was one in which a player either retained her behavior and chosen marker from the previous period or changed both. An unlinked choice was when she changed her behavior or marker but not both. A strong preference toward linked choices was present in general (Fig. 1), but it was significantly stronger in the specific case when a player received the optimal coordination payoff in the previous period (conditional logit, $P < 0.001$, table S3). These linked choices consisted almost exclusively of choices in which the player retained her behavior and marker from the previous period (figs. S1 and S2). In addition, although the preference for linked choices after coordinating on the optimal behavior was present in both treatments, it was significantly stronger in the marker-maintained treatment (conditional logit, $P < 0.001$, table S3). These results indicate that players showed a general tendency to couple behaviors and markers. This tendency, however, was strongest when a player hit upon a successful behavior-marker combination, and it was further reinforced and amplified in the marker-maintained treatment when the marker was not prevented from acquiring meaning.

Substantial linkage at the individual level produces covariation between behavior and marker at the aggregate level. If strong enough and specific enough, the linkage exhibited in the experiment should have produced a significant increase in covariation in the marker-maintained treatment, when it was possible, but not in the marker-randomized treatment, when it was not. Even though linkage was present, however, covariation should have been similar in the two treatments at the beginning of the experiment, before covariation had time to accumulate. Only in later periods should the covariation have been significantly higher in the marker-maintained treatment. The aggregate covariation between behavior and marker indeed followed this dynamical pattern. During the first five periods, the covariation was not different in the two treatments (Welch two-sample t test, df = 7.01, two-sided $P = 0.68$, Fig. 2A), whereas in the final five periods, the covariation was significantly higher in the marker-maintained case (Welch two-sample t test, df = 12.107, two-sided $P < 0.001$, Fig. 2A). Covariation thus strongly and significantly increased in the marker-maintained treatment but not in the marker-randomized case. This led to a strong overall treatment difference in the accumulation of the markers' predictive power [Newey-West regression (40), $P < 0.001$, Fig. 2A].

The presence of covariation does not mean that players will exploit it by assorting into groups characterized by shared markers. Players could simply fail to recognize the association between behavior and marker as it developed, or they could fail to recognize its usefulness. Nonetheless, players exhibited an increasing inclination to request partners with the same shape as covariation accumulated. Throughout the marker-randomized treat-

ment, players requested same-shape partners roughly 50% of the time (Fig. 2B), a result consistent with indifference concerning the two interaction policies. In the marker-maintained treatment, however, players increasingly requested partners having the same shape as time passed. This increase was highly significant (Newey-West regression, $P < 0.001$), and the vast majority of players (87%) requested partners with the same shape in the final five periods (Fig. 2B), indicating that ingroup favoritism became an almost universal phenomenon.

In the presence of covariation, this kind of ingroup favoritism should lead to more coordination and improved payoffs, but the strength of the effect will vary with the degree of covariation and preferential assortment. A calculation of the mean payoff over periods for each subject shows that payoffs were significantly different across the two treatments. The mean payoff in the marker-randomized treatment was 20.819 points, and it was 27.454 in the marker-maintained treatment (Welch two-sample t-test, df = 88.912, two-sided $P < 0.001$). This difference, however, depended specifically on the dynamical increase in the markers' predictive content in the marker-maintained treatment, and this fact is central to our finding that the evolution of cultural groups changed the incentives associated with ingroup favoritism. Specifically, for those players who requested a partner with the same shape, the mean payoff per period was not significantly different between the two treatments in the first five periods (Welch two-sample t-test, df = 123.139, two-sided $P = 0.1638$), whereas it was highly significant in the final five periods (Welch two-sample t test, df = 105.733, two-sided $P < 0.001$). The higher overall payoffs in the marker-maintained treatment stemmed from an increase in coordinating on the optimal behavior in each of the two subpopulations (Fig. 3). A detailed analysis formally confirms the substantial and robust payoff effect that resulted from assorting on markers in the marker-maintained treatment (37).

These results show how the evolution of cultural groups can reconstitute the social environment and produce selection for an ingroup bias that was not initially advantageous. If selective pressures of this sort were common in past human societies, a plausible outcome would arguably be a relatively inflexible bias leading individuals to prefer others similar in some symbolic dimension. This idea is consistent with much research showing an astonishing willingness for subjects to exhibit ingroup favoritism when groups are based on trivial, short-lived distinctions (12, 15, 16, 21, 22). For our study, this could mean that the marker-based assortment we documented largely reflected a readiness to favor the ingroup that was already in place when the subjects came to the lab, and it did not stem from the endogenous formation of cultural groups during the experiment. In particular, although we found a pronounced difference in assortment dynamics in our two treatments, we still found a strong tendency to assort in the marker-randomized treatment. This assortment was relatively meaningless with respect to payoffs, but

because requesting a partner with the same shape was free, it is consistent with two different motives on the part of players. Players could have simply been indifferent between two largely meaningless, cost-free social interaction policies, or they could have had a strong residual taste for same-shape partners even when such pairings did not improve payoffs. To distinguish between these two possibilities, we conducted a second experiment with three treatments, all of which required subjects to pay a small cost for ingroup favoritism.

In the three treatments of our second experiment (37), subjects had to pay a cost of 1 point when they requested and were successfully paired with a partner having the same shape. To maximize the salience of the marker, all players retained their chosen markers in all treatments. In the payoff-equivalent treatment, the payoff structure was changed such that coordinating on A or B yielded the same payoff (21 points) regardless of the players' subpopulation. Because the payoff structure in the subpopulations was identical in this case, players did not differ in terms of some unobservable variable related to payoffs, and thus they had no material problem the markers could help them solve. They could, of course, continue to bias their interactions toward those having the same marker if willing to pay the cost. In the fixed-marker treatment, players only chose a marker in the first period. This marker was then retained for all 80 periods. In this treatment, markers were ostensibly similar to traits like race that are often perceived as immutable. Because of this perceived immutability, which may or may not be an accurate perception, such traits are especially prone to essentialist generalizations and are thus prime candidates for generating ingroup favoritism and outgroup hostility (41). A truly immutable marker, however, like the one we implemented, should not evolve to be a stable predictor of behavior because individuals cannot adjust their markers to reflect changing social circumstances. In the fixed-marker treatment, for example, players could benefit from changing their expectations about where to coordinate when they changed subpopulations, but they could not change their markers to signal their shifting expectations. Players could nonetheless choose to assort on marker, if they wished. Lastly, as a new baseline, the modified marker-maintained treatment was similar to the original marker-maintained treatment, but it involved the same assortment cost used in the payoff-equivalent and fixed-marker cases.

As in the original marker-maintained treatment, the covariance between behavior and marker accumulated at a significant rate through time in the modified marker-maintained treatment [Newey-West (40) regression, maximal lag of 10, period trend t test, $P = 0.003$]. In early periods, the covariance in the fixed-marker treatment was lower than it was in the modified marker-maintained treatment, and this difference was marginally significant (z test on Newey-West estimated intercepts, $P = 0.067$). Furthermore, unlike the modified marker-maintained case, covariation did not accu-

mulate through time in the fixed-marker treatment (Newey-West regression, lag of 10, period trend t-test, $P = 0.294$). The estimated time trend was slightly negative, and this was significantly different from the positive trend in the modified marker-maintained treatment (z test on Newey-West estimated period coefficients, $P = 0.002$). In the payoff-equivalent treatment, covariance was significantly lower in early periods than it was in the modified marker-maintained treatment (z-test on Newey-West estimated intercepts, $P < 0.001$), and it declined even further at a significant rate (Newey-West regression, lag of 10, period trend t test, $P < 0.001$). In this case, covariance actually declined all the way to 0 because all players soon converged on A in all subpopulations. With no variation in behavior, covariation between behavior and marker is not possible. Shared history was sufficient to form accurate expectations about where to coordinate, and the marker was not useful in this respect. In sum, trivial groups became cultural groups in the modified marker-maintained treatment, but trivial groups remained trivial in the payoff-equivalent and fixed-marker treatments.

Players, in turn, responded strongly to the resulting variation in the accumulated predictive power of markers. In the modified marker-maintained baseline, roughly 55 to 60% of the players requested partners with the same shape in later periods (Fig. 4). In the payoff-equivalent and fixed-marker treatments, however, only 15 to 25% assorted on shape in later periods, and the differences relative to the baseline were highly significant (Fig. 4). The payoff-equivalent case is especially clear because, as mentioned above, all players eventually played A in both subpopulations, and the predictive value of the markers went to zero as a result. Correspondingly, the proportion of players requesting same-shape partners unraveled relentlessly as the experiment progressed (Fig. 4). The fact that assortment did not disappear altogether suggests that perhaps a few players had a weak taste for assortment even when this did not improve coordination. Altogether, however, our results show that the preference for interacting with similarly marked players varied strongly according to whether markers became accurate predictors of behavior in the face of heterogeneous behavioral expectations. In short, ingroup favoritism had little to do with an unconditional preference for similarly marked partners and a lot to do with whether trivial groups evolved into cultural groups. For this cultural evolutionary transition to happen, two requirements had to be met. First, players had to differ persistently in some important but unobservable dimension that could sustain symbolic representation. Our payoff-equivalent treatment removed this feature, and assorting on shape steadily declined through time. Second, the symbolic markers themselves had to be freely chosen and mutable in a way that allowed an association between markers and unobservables to develop. Our marker-randomized and fixed-marker treatments removed this feature, and assorting on shape was relatively low in all periods when

compared to their respective marker-maintained treatments.

The research on intergroup processes indicates that people have a willingness to show ingroup favoritism, and in particular this holds even when groups are trivial and evanescent (12, 13, 16–18, 20–22, 25). This research tradition has generally examined neither the evolutionary mechanisms behind group formation nor the impact of these mechanisms on ingroup favoritism. We implemented an experiment in which the significance of groups had to arise, if at all, endogenously, thus providing an evolutionary foundation for ingroup favoritism. In this setting, trivial groups remained trivial under certain circumstances, but under other circumstances they developed into cultural groups composed of individuals who shared both behavioral expectations and symbolic markers signaling group affiliation. Ingroup favoritism was strongly associated with cultural groups but not with trivial groups. Our experiments made exclusive use of coordination games, which serve as a kind of generic proxy for strategic settings with multiple equilibria. Many strategic settings are characterized by multiple equilibria (42), and thus the dynamical processes examined here have potentially broad significance. The mechanisms implicated in the evolution of human prosociality, for example, often produce multiple equilibria (43, 44), and so cooperation is a behavioral domain with considerable scope for the path-dependent evolution of groups with different norms and expectations. In this sense, cooperation can be analogous to coordination. Even more generally, whenever people have a shared interest in distinguishing among themselves in terms of their unobservable information (38), whatever that means in a given situation, the logic behind the evolution of cultural groups holds.

References and Notes
1. F. Barth, Ed., *Ethnic Groups and Boundaries: The Social Organization of Cultural Difference* (Little, Brown, Boston, 1969).
2. D. E. Blom, *J. Anthropol. Archaeol.* **24**, 1 (2005).
3. J. B. Eicher, Ed., *Dress and Ethnicity: Change Across Space and Time* (Berg, Oxford, 1995).
4. C. Silverman, *J. Am. Folkl.* **101**, 261 (1988).
5. C. Torres-Rouff, *Curr. Anthropol.* **43**, 163 (2002).
6. S. Bowles, *Microeconomics: Behavior, Institutions, and Evolution* (Russell Sage, New York, 2004).
7. T. Schelling, *Micromotives and Macrobehavior* (Norton, New York, 1978).
8. H. P. Young, *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions* (Princeton Univ. Press, Princeton, NJ, 1998).
9. R. L. Axtell, J. M. Epstein, H. P. Young, in *Social Dynamics*, S. N. Durlauf, H. P. Young, Eds. (MIT Press, Cambridge, MA, 2001), pp. 191–211.
10. S. Bowles, G. Loury, R. Sethi, "Is equal opportunity enough? A theory of persistent group inequality," Santa Fe Institute Working Paper; www.santafe.edu/~bowles.
11. H. Bernhard, U. Fischbacher, E. Fehr, *Nature* **442**, 912 (2006).
12. J. F. Dovidio, P. Glick, L. A. Rudman, Eds., *On the Nature of Prejudice: Fifty Years After Allport* (Blackwell, Oxford, 2005).
13. R. H. Fazio, J. R. Jackson, B. C. Dunton, C. J. Williams, *J. Pers. Soc. Psychol.* **69**, 1013 (1995).
14. E. L. Glaeser, D. I. Laibson, J. A. Scheinkman, C. L. Soutter, *Q. J. Econ.* **115**, 811 (2000).
15. L. Goette, D. Huffman, S. Meier, *Am. Econ. Rev.* **96**, 212 (2006).

16. M. Hewstone, M. Rubin, H. Willis, *Annu. Rev. Psychol.* **53**, 575 (2002).
17. K. D. Kinzler, E. Dupoux, E. S. Spelke, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12577 (2007).
18. Z. Kunda, *Social Cognition: Making Sense of People* (MIT Press, Cambridge, MA, 1999).
19. T. Yamagishi, N. Jin, T. Kiyonari, *Adv. Group Process.* **16**, 161 (1999).
20. M. B. Brewer, *Pers. Soc. Psychol. Bull.* **17**, 475 (1991).
21. M. Sherif, O. J. Harvey, B. J. White, W. R. Hood, C. W. Sherif, *Intergroup Conflict and Cooperation: The Robbers Cave Experiment* (Institute of Group Relations, Univ. of Oklahoma, Norman, OK, 1961).
22. H. Tajfel, M. G. Billig, R. P. Bundy, C. Flament, *Eur. J. Soc. Psychol.* **1**, 149 (1971).
23. R. Boyd, H. Gintis, S. Bowles, P. J. Richerson, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3531 (2003).
24. J.-K. Choi, S. Bowles, *Science* **318**, 636 (2007).
25. J. Correll, G. R. Urland, T. A. Ito, *J. Exp. Soc. Psychol.* **42**, 120 (2006).
26. R. Boyd, P. J. Richerson, *J. Theor. Biol.* **215**, 287 (2002).
27. R. Boyd, P. J. Richerson, *The Origin and Evolution of Cultures* (Oxford Univ. Press, Oxford, 2005).
28. R. Boyd, P. J. Richerson, *Culture and the Evolutionary Process* (Univ. of Chicago Press, Chicago, 1985).
29. W. H. Durham, Ed., *Coevolution: Genes, Culture, and Human Diversity* (Stanford Univ. Press, Stanford, CA, 1991).
30. F. J. Olding-Smee, K. N. Laland, M. W. Feldman, *Niche Construction: The Neglected Process in Evolution* (Princeton Univ. Press, Princeton, NJ, 2003).
31. P. J. Richerson, R. Boyd, *Not By Genes Alone: How Culture Transformed the Evolutionary Process* (Univ. of Chicago Press, Chicago, 2005).
32. It is not crucial for our purposes that identical actions lead to high payoffs. Instead, the necessity of coordinated choices for a high payoff is important. Thus, if the choice combination A for player 1 and B for player 2 led to a high payoff, then the players would simply need to coordinate on (A, B). Almost every bargaining problem or economic exchange involves some necessity for coordinating expectations and actions.
33. J. G. Jorgensen, *Western Indians* (Freeman, New York, 1980).
34. C. M. Judd, B. Park, in *On the Nature of Prejudice: Fifty Years After Allport*, J. F. Dovidio, P. Glick, L. A. Rudman, Eds. (Blackwell, Oxford, 2005), pp. 123–138.
35. S. Bowles, *Science* **314**, 1569 (2006).
36. R. F. Heizer, in *Handbook of North American Indians: California*, R. F. Heizer, W. C. Sturtevant, Eds. (Smithsonian Institution, Washington, DC, 1978), pp. 690–693.
37. Materials and methods are available as supporting material on *Science* Online.
38. R. Boyd, P. J. Richerson, *Cult. Anthropol.* **2**, 65 (1987).
39. R. McElreath, R. Boyd, P. J. Richerson, *Curr. Anthropol.* **44**, 122 (2003).
40. W. K. Newey, K. D. West, *Econometrica* **55**, 703 (1987).
41. J. T. Jost, D. L. Hamilton, in *On the Nature of Prejudice: Fifty Years After Allport*, J. F. Dovidio, P. Glick, L. A. Rudman, Eds. (Blackwell, Oxford, 2005), pp. 208–224.
42. D. Fudenberg, E. Maskin, *Econometrica* **54**, 533 (1986).
43. R. Boyd, P. J. Richerson, *Ethol. Sociobiol.* **13**, 171 (1992).
44. K. Panchanathan, R. Boyd, *Nature* **432**, 499 (2004).
45. This research was supported by the Swiss National Science Foundation (105312-114107) and is part of the Research Priority Program "Foundations of Human Social Behavior—Altruism versus Egoism" at the University of Zurich. We thank S. Bowles for valuable comments on an earlier version of this article and R. McElreath for helpful insights during the initial stages of the project.

Understanding Overbidding: Using the Neural Circuitry of Reward to Design Economic Auctions

Mauricio R. Delgado,[1] Andrew Schotter,[2] Erkut Y. Ozbay,[3] Elizabeth A. Phelps[4]*

We take advantage of our knowledge of the neural circuitry of reward to investigate a puzzling economic phenomenon: Why do people overbid in auctions? Using functional magnetic resonance imaging (fMRI), we observed that the social competition inherent in an auction results in a more pronounced blood oxygen level–dependent (BOLD) response to loss in the striatum, with greater overbidding correlated with the magnitude of this response. Leveraging these neuroimaging results, we design a behavioral experiment that demonstrates that framing an experimental auction to emphasize loss increases overbidding. These results highlight a role for the contemplation of loss in understanding the tendency to bid "too high." Current economic theories suggest overbidding may result from either "joy of winning" or risk aversion. By combining neuroeconomic and behavioral economic techniques, we find that another factor, namely loss contemplation in a social context, may mediate overbidding in auctions.

An unresolved question in the emerging field of neuroeconomics is whether data from neuroscience can inform economic theory such that it motivates behavioral economic institutional design (*1–4*). In this report, we address this question by taking advantage of our knowledge of the neural circuitry of reward to investigate a puzzling economic phenomenon. Specifically, why do people overbid in auctions? (*5, 6*).

Auctions are an old and widely used method in allocating goods (*7*). Mention of them dates back to Roman times, when spoils of war were sold on the block. Although there are many different types of auctions, they all share the feature that bidders must determine a bidding strategy (or bid function) to be used in submitting their bid. A bid function for a buyer in an auction is a mapping from the value that the bidder places on the good for sale to the bid chosen. A set of bidding functions is considered to be an equilibrium (Nash equilibrium) if, given the strategy used by one's opponents, no bidder has any incentive to change his or her bidding strategy. One robust finding in experimental auctions is that bidders tend to bid above their Nash equilibrium risk-neutral bid function (*5*); this behavior has been labeled "overbidding" in the economics literature. In other words, given the value of the good for sale they submit bids that are "too high." Two competing explanations for this phenomenon exist. Many scholars have assumed that risk aversion is responsible for this increase in bids, because bidding above one's risk-neutral Nash equilibrium bid function is exactly what risk aversion prescribes

(*5, 6, 8*). Another explanation stems from the ideas that bidders enjoy a "joy of winning" the social competition inherent in an auction (*5, 6*).

The goal of this study is to provide insight into the neural circuitry of experimental auctions and to use this insight to generate and test a behavioral economic approach to understand overbidding. First, we used functional magnetic resonance imaging (fMRI) to examine the neural correlates of winning and losing an experimental auction, while modulating potentially important variables such as type of social competition (auction versus lottery) and type of incentive (money versus points with no monetary value). On the basis of these brain imaging results and our understanding of the neural circuitry of reward, we generated a hypothesis concerning the mechanisms underlying overbidding in experimental auctions. We then tested this hypothesis in a behavioral economic experiment.

In the fMRI study, 17 participants were instructed that they would each be playing two types of games: a two-person auction and a lottery (52 events for each treatment) (*9*). Before participants were scanned, they briefly met their competitor for the auction and were informed that they would be playing an unknown but fixed strategy. In the auction game, participants were assigned a value (*V*) at the beginning of each trial. These values were drawn from a finite set with equal probability. Participants were asked to choose a bid (*b*) (the decision phase) and were then informed if they won or lost the auction (the outcome phase). There were four possible *V*'s assigned for the good sold (6, 8, 10, 12) and four options for *b* (2, 5, 7, 8). The competitor bid according to the Nash equilibrium strategy (*V:b* equals 6:2, 8:5, 10:7, 12:8). In the money condition, *V* and *b* represented dollars, and the participants were informed they would receive a payoff of *V* minus *b* if they won that trial and zero if they lost. They would be paid their total winnings from one randomly selected block out of the four

[1]Department of Psychology, Rutgers University, Newark, NJ 07102, USA. [2]Department of Economics, New York University, and Center for Experimental Social Science, New York, NY 10003, USA. [3]Department of Economics, University of Maryland, College Park, MD 20742, USA. [4]Department of Psychology, New York University, New York, NY 10003, USA.

*To whom correspondence should be addressed. E-mail: liz.phelps@nyu.edu