Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap

Geoffrey L. Cohen, *et al.*

Science **324**, 400 (2009);
DOI: 10.1126/science.1170769

28. M. Tranter, M. Skidmore, J. Wadham, *Hydrol. Process.* **19**, 995 (2005).
29. J. C. Priscu *et al.*, *Science* **286**, 2141 (1999).
30. J. L. Kirschvink, in *The Proterozoic Biosphere*, J. W. Schopf, C. Klein, Eds. (Cambridge Univ. Press, Cambridge, 1992), pp. 51–52.
31. P. F. Hoffman, D. P. Schrag, *Terra Nova* **14**, 129 (2002).
32. D. E. Canfield *et al.*, *Science* **321**, 949 (2008); published online 16 July 2008 (10.1126/science.1154499).
33. J. A. Mikucki, thesis, Montana State University (2005).
34. G. Eischeld and S. Fawcett assisted with oxygen isotope measurements, A. Masterson assisted with sulfur isotope measurements, and G. Gordon assisted with iron isotope measurements. We are grateful for assistance from the Crary Laboratory, Petroleum Helicopters Incorporated, and S. Carter. We would like to extend a special thanks to McMurdo personnel B. Peace and W. Hayworth. Discussions with H. Keys, M. Tranter, K. Welch, W. B. Lyons, W. Hamilton, F. MacDonald, S. Shah, and A. Knoll were helpful in formulating the ideas presented. The comments of three anonymous reviewers and A. J. Kaufman greatly improved the manuscript. This research was funded by an NSF Polar postdoctoral fellowship (OPP-0528710) to J.A.M. Additional support was provided by NSF grants EAR-0311937 (A.P.), OPP-432595 and OPP-0631494 (J.C.P.), and OPP-0338097 and OCE-0728683 (awarded to G. R. DiTullio) (P.A.L.); Canadian Institute for Advanced Research (A.V.T.); Harvard–Microbial Sciences Initiative; and NASA (NNX07AV51G) (D.T.J.). DNA sequences have been submitted to GenBank with accession numbers FJ389341 to FJ389351.

Supporting Online Material
www.sciencemag.org/cgi/content/full/324/5925/397/DC1
Materials and Methods
Fig. S1
Tables S1 to S4

17 October 2008; accepted 5 March 2009
10.1126/science.1167350

Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap

Geoffrey L. Cohen,[1]* Julio Garcia,[1] Valerie Purdie-Vaughns,[2] Nancy Apfel,[3] Patricia Brzustoski[3]

A 2-year follow-up of a randomized field experiment previously reported in *Science* is presented. A subtle intervention to lessen minority students' psychological threat related to being negatively stereotyped in school was tested in an experiment conducted three times with three independent cohorts (N = 133, 149, and 134). The intervention, a series of brief but structured writing assignments focusing students on a self-affirming value, reduced the racial achievement gap. Over 2 years, the grade point average (GPA) of African Americans was, on average, raised by 0.24 grade points. Low-achieving African Americans were particularly benefited. Their GPA improved, on average, 0.41 points, and their rate of remediation or grade repetition was less (5% versus 18%). Additionally, treated students' self-perceptions showed long-term benefits. Findings suggest that because initial psychological states and performance determine later outcomes by providing a baseline and initial trajectory for a recursive process, apparently small but early alterations in trajectory can have long-term effects. Implications for psychological theory and educational practice are discussed.

Whether and how psychological interventions produce lasting positive consequences are critical questions for scientists and policy-makers. This report presents evidence of how interventions, even brief or subtle, can produce lasting benefit when targeted at important psychological processes. It does so by focusing on the long-term impact of a psychological intervention designed to reduce the racial achievement gap through the lessening of academic underperformance.

The achievement gap between academically at-risk minority students and European American students has long concerned the educational community (*1*). In a society where economic success depends heavily on scholastic accomplishment, even partial remediation of this gap would be consequential. This is especially true for low-achieving students, given the societal, institutional, and personal costs of academic failure.

Research shows the importance of psychological factors in intellectual achievement (*2*–*4*). Situations where one could be judged or treated in light of a negative stereotype can be stressful and thus undermine performance (*5*–*7*). For African Americans in school, the concern that they or another African American could be seen as confirming a negative stereotype about their group's intelligence can give rise to stress and depress performance (*5*–*8*).

Findings of two randomized field experiments addressing this psychological threat in the classroom were reported in *Science* (*8*). These tested a values-affirmation intervention. Beginning early in seventh grade, students reflected on an important personal value, such as relationships with friends and family or musical interests, in a series of structured writing assignments. Such self-affirmations reduce psychological threat and stress (*9*–*11*) and can thus improve performance. The intervention should benefit students from groups subjected to threat pervasive enough to undermine their average performance—in this case, negatively stereotyped minority students. As predicted, relative both to a control group and to historical norms, one or two administrations of the intervention improved the fall-term grades of African Americans and lowered the psychological availability of the stereotype. European Americans were unaffected (*8*).

A 2-year follow-up is now reported. We assess whether the affirmation buffers minority students from the effects of psychological threat over the long term, leading to academic benefits beyond the short-term ones of a single academic term previously found. Generally, psychological processes and their consequences are examined for relatively brief periods, often in experimental studies lasting 30 min or an hour. By contrast, because the present study spans 2 years, its findings speak to how an apparently brief psychological intervention triggers processes that affect performance and psychological outcomes over considerable periods of time. Given the multitude of factors that could mute the effects of such processes in the classroom, the findings address the longevity and real-world significance of these processes. This is particularly important given that the effects of interventions and psychological manipulations often decay and may even reverse over time for reasons that are little understood (*12*, *13*).

Because chronic evaluation is a key aspect of school and work environments, performance in these settings can be self-reinforcing. A recursive cycle, where psychological threat lowers performance, increasing threat and lowering performance further, in a repeating process, can magnify early performance differences among students (*14*). Early outcomes set the starting point and initial trajectory of a recursive cycle and so can have disproportional influence. For instance, the low self-confidence of students who experience early failure, even by chance, is surprisingly difficult to undo (*15*). A well-timed intervention could provide appreciable long-term performance benefits through early interruption of a recursive cycle.

Results encompass the original two student cohorts and a third cohort run after the original two experiments. The cohorts were observed for a period running from the first term of seventh grade to the end of eighth grade, typically covering ages 12 to 14. Although the period involves the last 2 years of middle school, for clarity these will henceforth be referred to as Year 1 and 2, respectively. Individual students were randomly assigned to the affirmation condition or the control condition. The former completed affirmation exercises, the latter neutral exercises. The treatment consisted of variations on the original affirmation exercise in which students wrote about the personal importance of a self-defining value (*16*). The control exercises consisted of variations on the original control exercise in which students wrote about an unimportant value or a similarly

[1]Department of Psychology, University of Colorado, Muenzinger Psychology Building, Boulder, CO 80309–0345, USA. [2]Department of Psychology, Columbia University, 405 Schermerhorn Hall, New York, NY 10027, USA. [3]Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06520, USA.

*To whom correspondence should be addressed. E-mail: cohen.geoff@gmail.com

neutral topic like their morning routine. The experimental manipulation, given three to five times in the seventh grade, occurred at roughly equal intervals throughout the year.

With the exception of a treatment dosage manipulation introduced at the beginning of Year 2, all original condition assignments were preserved (*16*). At the start of Year 2, 50% of the affirmed students were randomly assigned to a booster condition. These students received between two and four additional affirmations in Year 2. All remaining participants completed control exercises. This would determine whether long-term intervention effects, if present, rested on the treatment's continued administration or were triggered by its early effects.

The key outcome was grade point average (GPA) in core academic courses (science, social studies, math, and English), as the intervention was administered in different courses over 2 years and its effect was found across core courses in the original studies (*8*). To increase statistical power, we combined data from the three cohorts, because the intervention's effect, if found, was expected to be small and was found to be consistent across cohorts (*16*).

Complete GPA data for 2 years were obtained for 93% of the original participants (N = 385). Attrition did not vary by experimental condition either overall or within racial group (*16*). Degrees of freedom are greater for earlier outcomes because of attrition. Multiple regression tested treatment effects (*16*). A positive effect of affirmation on average 2-year GPA emerged for African Americans but not for European Americans. As with short-term grades, a group × experimental condition interaction emerged for the new long-term data [B = 0.33, $t(321)$ = 3.59, $P < 0.001$] (table S1). African Americans earned a higher 2-year GPA in the affirmation condition than in the control condition [B = 0.24, $t(144)$ = 3.45, P = 0.001]. No treatment effect was found for European Americans [B = −0.07, $t(170)$ = −1.19, P = 0.236]. The treatment effect for African Americans emerged for GPA in both outcome years. The group × treatment interaction and treatment effect for African Americans was significant for each year [Year 1: interaction B = 0.25, $t(344)$ = 2.73, P = 0.007, treatment B = 0.18, $t(162)$ = 2.69, P = 0.008; Year 2: interaction B = 0.39, $t(321)$ = 3.25, P = 0.001, treatment B = 0.27, $t(144)$ = 3.03, P = 0.003].

If the intervention interrupts a recursive process, its effects should be larger for initially low-achieving African Americans, because low performance should trigger worsening performance. Affirmation should make their prior performance less predictive of subsequent achievement. A three-way interaction between racial group, condition, and a continuous measure of pre-intervention performance on average 2-year GPA shows this [B = −0.32, $t(319)$ = −2.59, P = 0.010] (*16*). A two-way interaction between condition and pre-intervention performance emerged for African Americans [B = −0.21, $t(144)$ = −2.49, P = 0.014], not European Americans [B = 0.10, $t(170)$ = 1.10, P = 0.274]. Regardless of previous performance level, European Americans were unaffected by the intervention. However, the affirmation effect was significant for low-performing African Americans, those at the 25th percentile of pre-intervention performance for their racial group [B = 0.41, $t(144)$ = 4.41, $P < 0.001$]. Although the affirmation effect was present in the first term for high-performing African Americans, those at the 75th percentile of pre-intervention performance for their group [Fig. 1; B = 0.19, $t(160)$ = 2.30, P = 0.019], it decayed and did not reach significance on 2-year GPA for them [B = 0.15, $t(144)$ = 1.67, P = 0.096]. At mean or moderate pre-intervention performance, treatment effects were virtually identical to those in the overall analysis (*16*).

Affirmed African Americans should be more likely to maintain their performance over time if the intervention interrupted a recursive process of worsening performance. Indeed, the downward trend in performance commonly found in middle school (*17*) was less steep for these students than for African Americans in the control condition, not just for one term but across 2 years. Although all children performed progressively worse with time (Fig. 1), the linear decline in annual GPA was smaller among affirmed than nonaffirmed African Americans [$F(1,146)$ = 7.36, P = 0.007] (*16*). The decline among European Americans did not vary by condition [$F(1,172)$ = 1.37, P = 0.24; group × condition × measure interaction, $F(1,323)$ = 7.41, P = 0.007]. Figure 2 illustrates how the performance trajectory of low-achieving African Americans angles upward after the intervention, keeping the gap between them and European Americans from widening with time.

Although the initial treatment had long-term performance effects, the dosage manipulation did not moderate the treatment effect on Year 2 GPA for either racial group or for any pre-intervention performance subgroup [| t's | < 1.3, P's > 0.20]. This further supports the presence of a recursive process, as the intervention's early effects suffice to explain its long-term effects (*16*). All students, including African Americans, tended to perform relatively worse in Year 2 if they had performed poorly in Year 1, even controlling for pre-intervention performance (*16*). That the treatment effect on Year 2 GPA was significantly mediated by Year 1 GPA suggests that this natural performance cycle could have carried forward the intervention's early impact (SOM Text).

The intervention's impact on students' psychological environment is indicated by data suggesting that it buffered African Americans against the impact of early poor performance on their long-term perceptions of adequacy. A survey assessed students' self-perceived ability to fit in and succeed in school—their adaptive adequacy in the academic environment (*16*). These data indicate

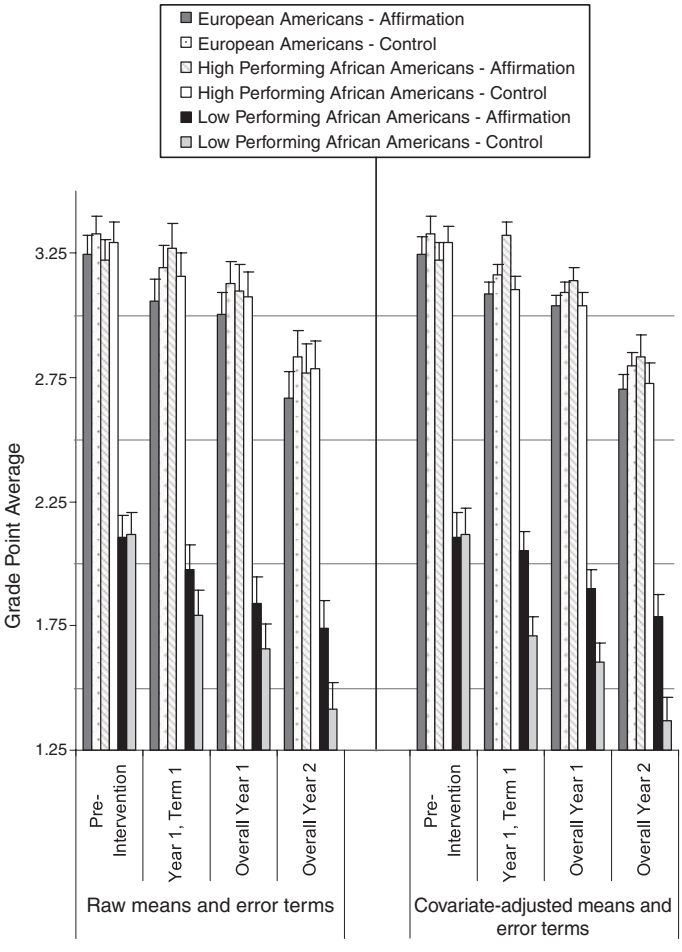

Fig. 1. Mean GPA in core courses, as a function of student group (African American versus European American), experimental condition, and pre-intervention level of performance of African Americans (an average of the prior year's GPA and pre-intervention seventh-grade performance). Data from participants with complete data are presented. Error bars represent standard errors. African Americans were categorized into low and high performers based on a median split within their racial group, reflecting their relative standing within their group. Year 1, Term 1 represents the first term after the initiation of the intervention. **(Left)** Raw means and error terms. **(Right)** Means and error terms adjusted for baseline covariates and students' assigned teacher team. The scale reflects the grade metric, ranging from 0 (= F) to 4.33 (= A+).

that the intervention uncoupled African Americans' long-term perceptions of adequacy from early poor performance. African Americans who had performed poorly early in the school year, and then received the affirmation, maintained a sense of their ability to fit and succeed in school over time. They had similar levels of self-perceived adequacy at the beginning and end of the year [paired $|t| < 0.2$]. For them, as for European Americans, early poor performance bore little relationship to their perceptions of adequacy at year's end, controlling for baseline perceptions [B's < 0.04, $|t$'s$| < 1$]. By contrast, for African Americans in the control condition, performing poorly before the manipulation predicted more negative perceptions of adequacy later [$B = 0.23$, $t(155) = 3.79$, $P < 0.001$]. They had lower self-perceived adequacy at the end of the year than they had had at the beginning [paired $t(40) = -2.45$, $P = 0.019$]. Low-performing African Americans thus ended the year with a lower sense of personal adequacy in the control condition than in the affirmation condition [$B = 0.31$, $t(155) = 3.30$, $P = 0.001$], with the latter not differing from European Americans [$|t| < 1$]. A mid-year assessment, which due to pragmatic constraints involved a shorter scale and only the first two cohorts, yielded the same results. Without intervention, early poor performance for minority students appeared to deliver a lasting blow to their sense of adequacy (*18*).

Although end-of-year adequacy correlated with higher GPA [$R = 0.23$, $P < 0.001$], statistical evidence that it mediated the treatment effect on GPA was not found (*16*). This suggests that the intervention might have discrete effects on a host of education-relevant psychological and behavioral outcomes. Here the intervention weakened the relationship not only between past and future performance, but also between past performance and later psychological state.

We also explored the effect of the intervention on students' assignment by their school to two major performance tracks—whether students were placed in remediation (assigned to a remedial program or held back in their grade), and whether they received advanced placement in math (*16*). Of the 13 students in the sample placed in remediation after the intervention, 11 were in the control condition (6%, versus 1% in the affirmation condition). Because counts for European Americans receiving the intervention were zero, we tested main effects of affirmation and racial group separately (*16*) (fig. S1). Logistic regression yielded a condition effect, with fewer affirmation-treated students placed in remediation [$\Delta\chi^2$ (1) = 14.06, $P < 0.001$]. Additionally, fewer European Americans (2%) were placed in remediation than African Americans (6%) [χ^2 (1) = 4.03, $P = 0.045$]. However, fewer affirmed African Americans were so classified than nonaffirmed African Americans [3% versus 9%; $\Delta\chi^2$ (1) = 9.31, $P = 0.002$]. This condition effect was confined to previously low-performing African Americans [5% versus 18%] (*16*). Condition effects were virtually identical in a rare events logistic regression (*19*).

Evidence of a positive treatment effect regarding assignment to advanced placement in math was found for African Americans (SOM Text) (*16*).

A values-affirmation intervention closed the achievement gap not only over one school term, but throughout African Americans' tenure in middle school. It also decreased the number of African Americans identified as at-risk and enrolled in remediation. Moreover, the intervention benefited those most in need and often least affected by traditional intervention—low-achieving students (*20*).

In chronically evaluative settings such as school, performance issues from self-reinforcing or recursive processes. A feedback loop, with psychological threat and poor performance reinforcing one another, can create worsening performance over time. Students' poor performance may also cause them to be seen as less able by their teachers and less worthy of attention and mentoring, increasing the likelihood of lower performance (*21*). The ability of the intervention to interact with recursive processes lies at the heart of how its effects persisted for 2 years. Because initial psychological states and early performance establish the starting point and initial trajectory of a recursive cycle, they can have disproportionate influence on long-term outcomes. When such recursive cycles are interrupted early, baseline outcomes and the long-term performance trajectories following from them can be changed. That a new starting point and trajectory for the recursive cycle was introduced by the affirmation is suggested by its weakening of the relationship between early poor performance and later performance and felt adequacy.

The following findings provide evidence for the intervention's interruption of a recursive cycle. First, early poor performance was less predictive of later performance and psychological state for affirmed African Americans than for nonaffirmed ones, suggesting that the intervention reset the starting point of a recursive cycle. Second, the affirmation not only benefited GPA, but also lifted the angle of the performance trajectory and thus lessened the degree of downward trend in performance characteristic of a recursive cycle. Third, the affirmation's benefits were most evident among low-achieving African Americans. These are the children most undermined by the standard recursive cycle with its worsening of performance and magnifying of initial differences in performance. Fourth, the affirmation prevented the achievement gap from widening with time. Fifth, treatment boosters were not needed to sustain its impact into Year 2. This indicates that processes triggered by the intervention in Year 1 suffice to explain its effect in Year 2. That the intervention's first-year

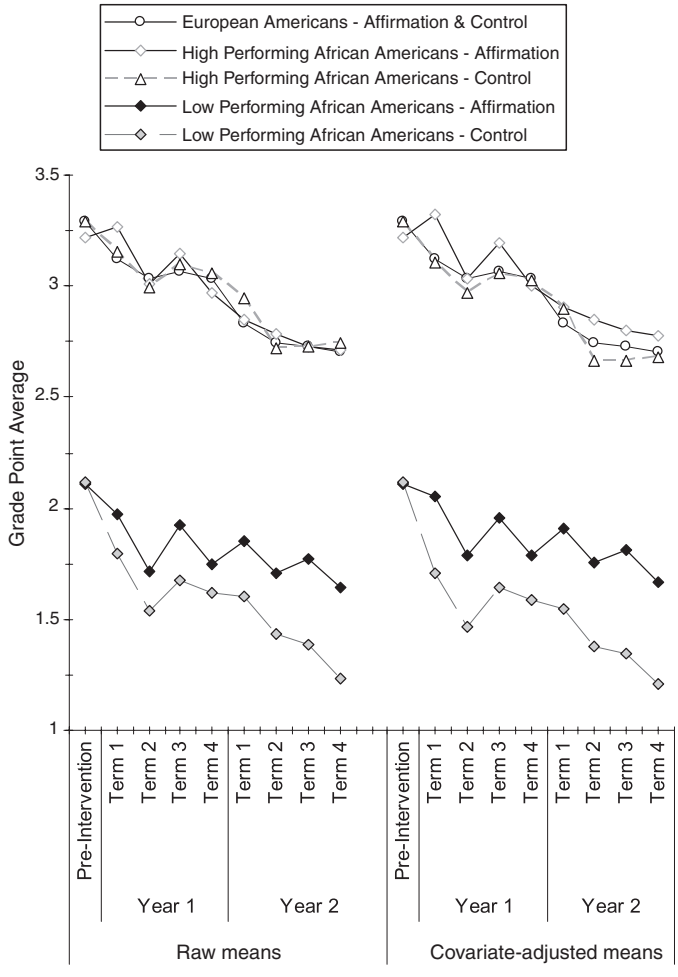

Fig. 2. Mean GPA in core courses for each term over 2 years, as a function of student group (African American versus European American), experimental condition, and pre-intervention level of performance of African Americans (an average of the prior year's GPA and pre-intervention seventh-grade performance). Data from participants with complete data are presented. African Americans were categorized into low and high performers based on a median split within their racial group, reflecting their relative standing within their group. Because European Americans in the two conditions did not differ significantly, their data were combined. **(Left)** Raw means. **(Right)** Means adjusted for baseline covariates and students' assigned teacher team.

impact mediated much of this effect further supports this notion.

Finally, students' psychological state sheds light on how affirmation processes interact with the recursive cycle. African Americans, a stereotyped group, displayed greater psychological vulnerability to early failure. For them, early failure may have confirmed that the stereotype was in play as a stable global indicator of their ability to thrive in school. By shoring up self-integrity at this time, the affirmation helped maintain their sense of adequacy and interrupted the cycle in which early poor performance influenced later performance and psychological state. Students' performance and psychological trajectory can be strongly influenced by timely actions, even when apparently small, that alter or reset the trajectory's starting point.

Other factors, such as teachers' expectancies of their students, could contribute to the longevity of the treatment's effect (*21*). For instance, that fewer affirmed children were assigned to remediation suggests that the intervention's effects were not only noted by the academic system, but acted upon by it.

The findings demonstrate how initial psychological processes, triggered by an apparently subtle intervention, can have psychological and pragmatic effects that perpetuate themselves over extended time spans, in the present case 2 years (*6*, *13*). They demonstrate the role of such processes in long-term intellectual achievement and also suggest a practical strategy for addressing the achievement gap. Effective psychological interventions depend on the presence of positive and sufficient structural, material, and human resources. Together with such resources and other educational programs, psychological interventions can help individuals perform to their potential and produce lasting positive changes in equity and opportunity.

References and Notes

1. C. Jencks, M. Phillips, *The Black-White Test Score Gap* (Brookings Institution, Washington, DC, 1998).
2. A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory* (Prentice Hall, Englewood Cliffs, NJ, 1986).
3. C. S. Dweck, C. Chiu, Y. Hong, *Psychol. Inq.* **6**, 267 (1995).
4. E. Zigler, E. C. Butterfield, *Child Dev.* **39**, 1 (1968).
5. C. M. Steele, S. J. Spencer, J. Aronson, in *Advances in Experimental Social Psychology,* M. P. Zanna, Ed. (Academic Press, San Diego, CA, 2002), pp. 379–440.
6. J. Aronson, C. B. Fried, C. Good, *J. Exp. Soc. Psychol.* **38**, 113 (2002).
7. G. L. Cohen, J. Garcia, *J. Pers. Soc. Psychol.* **89**, 566 (2005).
8. G. L. Cohen, J. Garcia, N. Apfel, A. Master, *Science* **313**, 1307 (2006).
9. J. D. Creswell *et al.*, *Psychol. Sci.* **16**, 846 (2005).
10. C. M. Steele, in *Advances in Experimental Social Psychology*, L. Berkowitz, Ed. (Academic Press, New York, 1988), pp. 261–302.
11. D. K. Sherman, G. L. Cohen, in *Advances in Experimental Social Psychology*, M. P. Zanna, Ed. (Academic Press, San Diego, CA, 2006), pp. 183–242.
12. R. Schulz, B. H. Hanusa, *J. Pers. Soc. Psychol.* **36**, 1194 (1978).
13. L. Ross, R. Nisbett, *The Person and the Situation* (McGraw-Hill, New York, 1991).
14. T. D. Wilson, P. W. Linville, *J. Pers. Soc. Psychol.* **49**, 287 (1985).
15. M. R. Lepper, L. Ross, R. Lau, *J. Pers. Soc. Psychol.* **50**, 482 (1986).
16. Materials and methods are available as supporting material on *Science* Online.
17. J. S. Eccles, S. Lord, C. Midgley, *Am. J. Educ.* **8**, 520 (1991).
18. For the adequacy outcome, the racial group × condition × prior performance interaction was significant [$B = -0.31$, $t(328) = -2.54$, $P = 0.011$], indicating that while there was no condition × prior performance interaction among European Americans [$B = 0.11$, [$|t| < 1.1$], there was such an interaction among African Americans [$B = -0.20$, $t(155) = -2.75$, $P = 0.007$].
19. M. Tomz, G. King, L. Zeng, RELOGIT: Rare events logistic regression, v. 1.1; Harvard Univ., Cambridge, MA, 1 October 1999, http://gking.harvard.edu/.
20. S. J. Ceci, P. B. Papierno, *Am. Psychol.* **60**, 149 (2005).
21. R. Rosenthal, *Curr. Dir. Psychol. Sci.* **3**, 176 (1994).

22. We thank the student participants and their parents, the teachers, staff, and administrators of the school district for their involvement in the project. We also thank E. Zigler, D. Green, C. Steele, E. Pronin, D. Sherman, G. Walton, J. Correll, C. Judd, J. Cook, E. Paluck, S. Taborsky-Barba, S. Tomassetti, and S. Wert for their help and feedback. This research was supported primarily by grants from the William T. Grant Foundation and the Russell Sage Foundation. Additional support was provided by the Nellie Mae Education Foundation and the Institute for Social and Policy Studies of Yale University.

Supporting Online Material

www.sciencemag.org/cgi/content/full/324/5925/400/DC1
SOM Text
Fig. S1
Table S1
References

12 January 2009; accepted 26 February 2009
10.1126/science.1170769

Mirror Neurons Differentially Encode the Peripersonal and Extrapersonal Space of Monkeys

Vittorio Caggiano,[1] Leonardo Fogassi,[2,3] Giacomo Rizzolatti,[3] Peter Thier,[1] Antonino Casile[1]*

Actions performed by others may have different relevance for the observer, and thus lead to different behavioral responses, depending on the regions of space in which they are executed. We found that in rhesus monkeys, the premotor cortex neurons activated by both the execution and the observation of motor acts (mirror neurons) are differentially modulated by the location in space of the observed motor acts relative to the monkey, with about half of them preferring either the monkey's peripersonal or extrapersonal space. A portion of these spatially selective mirror neurons encode space according to a metric representation, whereas other neurons encode space in operational terms, changing their properties according to the possibility that the monkey will interact with the object. These results suggest that a set of mirror neurons encodes the observed motor acts not only for action understanding, but also to analyze such acts in terms of features that are relevant to generating appropriate behaviors.

Mirror neurons are a set of neurons, first described in the monkey premotor area F5, that respond both when the monkey performs an active goal-directed motor act and when he observes the same motor act performed by others (*1*, *2*). The most accepted interpretation of the function of mirror neurons is that they are involved in action understanding. Here, we investigated whether mirror neurons, besides playing a role in this function, also encode aspects of the observed actions that are relevant to subsequent interacting behaviors. A way to test this hypothesis is to examine the effect of relative distance between observer and actor on mirror neuron responses. Although completely irrelevant for "understanding" what the actor is doing, a precise knowledge of the distance at which the observed action is performed is crucial for selecting the most appropriate behavioral reaction.

To investigate quantitatively the possible degree of spatial modulation of the visual responses of mirror neurons, we first isolated hand movement–related neurons in area F5 of two rhesus monkeys by measuring the neurons' discharge while each monkey was executing hand goal-directed motor acts. The visual properties of these neurons were then assessed by having the experimenter perform the same motor acts in the monkey's peripersonal and extrapersonal (*3*–*7*) space, respectively (Fig. 1, A and B). The position of the experimenter's body was the same in all conditions, and actions were performed in the middle sagittal plane of the monkey's body. The selectivity for one of the two regions of space was then assessed by means of quantitative statistical analysis of the response patterns of 105 mirror neurons recorded from two monkeys (*8*).

Figure 2A shows the visual responses of three mirror neurons to motor acts executed in the peri- or extrapersonal space of the monkey. All three neurons responded during active movements of the monkey. However, their visual responses exhibited different types of tuning depending on whether the observed actions were executed in the monkey's peri- or extrapersonal space. Of all F5 mirror neurons tested, 26% ($n = 27$) exhibited a

[1]Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, University of Tübingen, 72076 Tübingen, Germany. [2]Dipartimento di Psicologia, Università di Parma, 43100 Parma, Italy. [3]Dipartimento di Neuroscienze e Istituto Italiano di Tecnologia, Università di Parma, 43100 Parma, Italy.

*To whom correspondence should be addressed. E-mail: antonino.casile@uni-tuebingen.de

www.sciencemag.org/cgi/content/full/324/5925/400/DC1

Supporting Online Material for

Recursive Processes in Self-Affirmation: Intervening to Close the Minority Achievement Gap

Geoffrey L. Cohen,* Julio Garcia, Valerie Purdie-Vaughns, Nancy Apfel, Patricia Brzustoski

*To whom correspondence should be addressed. E-mail: geoffrey.cohen@colorado.edu

Published 17 April 2009, *Science* **324**, 400 (2009)
DOI: 10.1126/science.1170769

This PDF file includes:

Supporting Online Material

Materials and Methods

Subjects

With the inclusion of the data from the new, third cohort of students, the following is based on the supporting material provided in our initial *Science* report (*S1*). Participants were recruited for the project via permission slips sent to students' parents through the mail and/or distributed in the classroom. These slips allowed parents to indicate whether they accepted or declined their child's participation in the study. Parents' acceptance included granting the researchers access to their child's official school grades. Of the total students enrolled, approximately 63% returned their permission slips in each year of the project. As access to special education students was limited at the start of the project, participants were predominately regular education students. Of those students returning permission slips, approximately 81% provided consent each year.

Of the original sample, 4 students in Experiment/Cohort 1, 34 students in Experiment/Cohort 2, and 19 students in Experiment/Cohort 3 were excluded. This occurred for the following reasons: absence at the time of the administration of the intervention (Experiment 1, n = 3; Experiment 2, n = 5); students' enrolled in a classroom not participating in the intervention (Experiment 2, n = 27; Experiment 3, n = 3); students' withdrawing from the school or being retained in 6th grade before the commencement of the study in 7th grade (Experiment 3, n = 15); missing data

(Experiment 1, n = 1); students' not understanding experimental instructions (Experiment 3, n = 1); and experimenter error (i.e., students received both treatment and control exercises; Experiment 2, n = 2).

The resulting baseline sample with performance data in Experiment 1 consisted of 133 students (60 males, 73 females). Of these, 38% were African American, 46% European American, 11% Hispanic or Latino American, and 5% Asian American. The final sample in Experiment 2 consisted of 149 students (76 males, 73 females). Of these, 46% were African American, 42% European American, 6% Hispanic or Latino American, 4% Asian American, and 1% "other." The final sample in Experiment 3 consisted of 134 students (63 males, 71 females). Of these, 42% were African American, 48% European Americans, 10% Hispanic or Latino American, and 1% Asian American.

As in our original *Science* report (*S1*), our analyses focused on African American students and European American students due to the small sample sizes of the other ethnic groups at the school site. However, virtually identical results are obtained if a bimodal category, comprised of Latino Americans and African Americans in a "negatively stereotyped" group (*S2*), and European Americans and Asian Americans in a "non-negatively stereotyped" group, is used.

Attrition. Complete data were obtained from 93% of the study participants, as 385 students of the 416 students who initially took part in the study provided complete data for each term over the two years of the study. Incomplete data from 31 students occurred for the following reasons: 29 students moved from the school district or left the school. Additionally, 2 students were excused

from the normal grading policies for at least one term due to extraneous factors such as medical issues. Of the 31 students, 8 left or did not provide grade data during Year 1, 7 in Year 2, and 16 left in the summer between the two years.

Attrition did not vary by experimental condition either across all participants [χ^2 (1) = 0.59, P = 0.443] or in the focal racial group of African Americans [χ^2 (1) = 0.01, P = 0.939]. However, attrition did vary by group, with African Americans having higher attrition rates (10%) than European Americans (4%) [χ^2 (1) = 6.09, P = 0.014].

Experimental tasks

Students participated in the study for two years. Due to pragmatic constraints, the frequency of the intervention varied somewhat from year to year, and cohort to cohort. For Experiment/Cohort 1, there were five administrations of the intervention in Year 1 (7th grade) and two in Year 2 (8th grade). For Experiment/Cohort 2, there were three administrations of the intervention in Year 1 and four in Year 2. For Experiment/Cohort 3, there were four administrations of the intervention in each year.

At each administration of the intervention, students received an envelope that was marked with their name. These envelopes contained exercises for either the control condition or the treatment condition. All teachers were provided with an identical script and procedure to follow in introducing and distributing the envelopes containing the exercise (e.g., “In class today, you’re

going to be doing a writing assignment for me"). They were also provided with specific answers to questions students might raise.

Teachers were kept blind to students' assignment to condition through a number of steps. Among these was the distribution of the treatment and control exercises in closed envelopes. Only students opened their envelopes, and upon completion of the exercise, they placed it back into the envelope and sealed it. Additionally, identical envelopes containing the exercises were distributed to all students in each teacher's class. Moreover, the exercises provided students with self-explanatory instructions, requiring virtually no guidance from teachers. Teachers were also instructed to remain at their desk while students independently and silently completed the exercise. We further minimized any potential teacher contamination effects by withholding from the teachers critical information about the nature of the exercises. This included, among other things, which exercise constituted the treatment and which the control, the differences in content between the exercises, and the hypothesized impact of the exercises. Additionally, the visual appearance of the treatment and control exercises was virtually identical in format and structure. While the differences in the content of the exercises lay in relatively small but important details, their visual appearance was virtually the same.

The written instructions used to guide students through the exercises had previously been thoroughly tested to ensure that they were intelligible, age-appropriate, and self-explanatory. Both affirmation and control exercises followed procedures similar to those developed and validated in prior research (*S3, S4*). In both conditions, subjects were presented with a short packet at each intervention administration.

Basic format of first interventions. With some minor variations, in each year the first two interventions followed the same basic format. The written instructions in the first intervention informed all subjects that they would be providing written responses to questions about "your ideas, your beliefs, and your life." The instructions further emphasized that while answering the various questions in the exercise, they should keep in mind that, "there are no right or wrong answers." The same set of values were listed on the cover page of the packet in both conditions: *athletic ability, being good at art, being smart or getting good grades, creativity, independence, living in the moment, membership in a social group (such as your community, racial group, or school club), music, politics, relationships with friends or family, religious values,* and *sense of humor.*

To provide a more difficult test of the hypothesis, the first intervention in Cohorts 2 and 3 excluded the value *being smart or getting good grades*. Previous research shows that there is flexibility in the sources of self-integrity, so that one's ability to endure threats in one domain can be bolstered by one's identity in a different domain (*S5, S6*). Given this, if people are protecting their sense of *global* worth or "self-integrity," the intervention's impact should be evident if they self-affirm in a domain different from the one in which they are threatened, in the present case academics.

For the first intervention, subjects in each condition were asked to read the list of values and to think about each one. For Cohort 1, subjects in both conditions were asked to mark the value "that is most important to you" with an "M" and the value "that is least important to you" with

an "L." For Cohorts 2 and 3, the task was simplified. In the treatment condition subjects were directed to circle their two or three *most* important values and in the control condition they were directed to circle their two or three *least* important values. For both experiments, students in each condition were informed that although several of the values might be important/not important to them, they should select only the requested number of values.

The next page of the packet directed subjects in the affirmation condition to "look at the value[s] you picked as most important to you," and to think about times when "this value" (Cohort1) or "these values" (Cohorts 2 and 3) were "important to you." They were then instructed to describe "in a few sentences" why their selected value/s were important to them. The following statement was included to reduce any evaluation apprehension that might otherwise be evoked: "Focus on your thoughts and feelings, and don't worry about spelling, grammar, or how well written it is." The instructions were virtually identical for subjects in the control condition, except that they instructed students to think about times when their *least important* value/s might be important to *someone else*, and to describe why the value/s might be important to someone else (*S3, S4*).

For the first intervention, the manipulation was reinforced on the final page of the packet. Students in the affirmation condition were asked to list the top two reasons why the value/s they had selected were important to them. Students in the control condition were asked to list the top two reasons why *someone else* would view the chosen value/s as important. Finally, to further increase the intervention's potential impact, subjects were asked to indicate their level of agreement with four easy-to-agree-with statements concerning their selections (*S7*). In the affirmation condition, examples included, "This value has [these values have] influenced my

life"; "In general, I try to live up to this value [these values]" and "This value is [These values are] an important part of who I am." The questions in the control condition, although similar, focused on other people (e.g., "These values have influenced some people"). Students indicated their response to each statement using separate scales that offered six response options, ranging from "strongly disagree" to "strongly agree."

Subsequent interventions. After the first two interventions in a given year, students completed a series of structured affirmation or control exercises over the remainder of the year. For the affirmation condition, these involved using different sets of values, changing the nature of the writing task (e.g., by having students write about why a certain value would be important to them over winter break), and personalizing the intervention by focusing it on a value each student had singled out in a previous affirmation exercise. For the control condition, these involved having students write about a daily routine, such as what they do when they get up in the morning or after they return home from school.

Performance outcomes

Grade point average (GPA). Academic performance was calculated using end-of-term official report cards provided by the school administration. The primary outcome was average GPA in core courses—science, social studies, math, and English/language arts. Students' GPA was assessed over 7th and 8th grade, with the exception of two students held back in the 7th grade, whose data in Year 2 came from their second year in that grade.

Enrollment in remediation. These data were also obtained from students' official school records. A composite score was created consisting of the three types of remediation indices provided by the school over the life of the project. These included: enrollment in a special school or tutoring program to help students catch up academically; classification of a child as exhibiting learning/emotional/behavioral difficulties interfering with school work, typically accompanied by enrollment in a special assistance program; and retention in either 7th or 8th grade, that is, being held back in the same grade.

Advanced placement in math. These data were also obtained from students' official records. Results are presented in the Supporting Text at the end of this document.

Psychological outcome of self-perceived adequacy

A survey assessing students' sense of adequacy in the academic environment—their ability to fit in and succeed in school—was administered once at the beginning of 7th grade (Year 1) prior to the experimental intervention, and then once at the end of the academic year. This scale was based on one used by the researchers in previous intervention research examining African American college students' sense of fit in school (*S8*). In order to make it appropriate for the younger age group in the present research, the scale was shortened and its language simplified. The survey consisted of two subscales. One subscale included 5 items to assess *self-perceived social belonging in school* (e.g., "People in my school accept me"; $\alpha = 0.77$). The second subscale included 4 items to assess *self-perceived ability to succeed in school* (e.g., "I know what I need to do to succeed at [school name]"; $\alpha = 0.66$). Together these two scales assess students'

self-perceived "goodness of fit" in the academic environment. Students responded to each item using a 6-point scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). The reliability of the entire scale was found to be adequate [$\alpha = 0.79$]. In order to further simplify the scale, for Cohort 3, one item with the lowest correlation with the overall scale mean [$R = 0.49$, $P < 0.001$] was removed. All but 11 students in the three cohorts completed the survey at time 1 and all but 13 completed it at time 2.

In order to assess the validity of the two subscales, a principal components analysis using varimax rotation was performed on the 9 scale items. In this analysis missing values were replaced by the item mean; virtually the same results are obtained if cases with missing values are excluded in the analysis. As expected, the analysis produced two discrete components. The first component pertained to *social belonging* (eigenvalue = 3.52), and the second component pertained to *ability to succeed* (eigenvalue = 1.23). The items comprising each of these components displayed absolute loading values ranging from 0.56 to 0.77, while at the same time registered no absolute loading value greater than 0.3 on a component other than its own. Pre-manipulation and post-manipulation composites were created for both components after reverse-coding negatively valenced items, and then averaged to provide mean ratings of self-perceived adequacy at both time points.

Data Analysis

Official course grades data. Analyses were conducted on participants providing complete data for the outcomes of interest. As in our previous report (*S1*), for the analyses of official grades

multiple regression was used to test the group x condition interaction and to compute the treatment effect for each of the two racial groups separately. A dummy variable was created for experimental condition (0 = control, 1 = treatment). Another dummy variable was created for student group (0 = European American, 1 = African American), and still another for student gender (0 = male, 1 = female). Two additional dummy variables were created to code for the three possible sets of teachers that each student could have in Year 1. These are referred to as teacher teams. Students belonging to the same team generally shared the same teachers for their core courses. For analyses of Year 2 GPA and overall GPA across both years, four dummy variables were created to code for the five possible teacher teams in Year 2. We controlled for Year 2 team because of the variability associated with different teams and teachers, their different curricula, and their different grading systems. However, these are not technically baseline variables, given that assignment to Year 2 team occurred after the initiation of the interventions. However, excluding these latter dummy variables from analyses does not change the results. Finally, two dummy variables coded for the three yearly cohorts of students.

As in our previous report (*S1*), the tested covariates included a standardized measure of baseline or pre-intervention performance in 7th grade, GPA in core courses taken in the previous year, and a standardized measure of pre-intervention state achievement test performance. All covariate data were obtained from teachers' official, pre-intervention gradebooks or official student transcripts. Two students had missing values for 7th grade pre-intervention performance, while a third student was missing previous year's GPA. To avoid losing these observations, a regression-estimated value based on the students' standing along the other baseline variables was used as substitute value for the former two students. This was advisable as these baseline variables could

account for much of the variance in the predictor; in particular previous year's GPA correlated highly [$R = 0.66$, $P < 0.001$]. In the latter student's case, academic grades from 5th grade were used. Results were not changed by simply dropping these participants from analyses, or by using any various alternatives methods for dealing with these missing values, such as substituting the sample mean for each missing value and then including in the models a dummy variable to code whether the value was missing or not. Pre-intervention state-wide test performance did not consistently account for unique variance in the analyses above and beyond the two other performance-based covariates and was thus excluded from the models. Also, as noted in the supporting material accompanying our original report, because the grading system used by each teacher was different, and thus the predictive power of baseline 7th grade performance could vary as a function of teacher, our models tested terms representing the interaction of this performance outcome with each of the relevant teacher team dummy variables (*S1*). However, these too did not account for unique variance in outcome and accordingly they were not retained in the models; retaining them did not change the results.

Regression analyses of Year 1 GPA used the Year 1 teacher team dummy variables. Regression analyses both of Year 2 GPA, and of overall GPA across Years 1 and 2, used team dummy variables from both Year 1 and Year 2. As noted previously, dropping the Year 2 team dummy variables does not change the results.

As in our previous report (*S1*), the 2-way interaction between student group (African American vs. European American) and experimental condition was computed in a full regression model that included main effects of student racial group, student gender, experimental condition,

teacher team, and the enumerated covariates, and all main effects and 2-way interactions involving student racial group, student gender, and experimental condition. Additionally, the two dummy variables coding for the three cohorts were also included. Table S1 presents the regression coefficients, standard errors, and *t* values for each term in the full regression model. In computing the treatment's effect on African Americans, we similarly controlled for the same specified covariates and main effects. This permitted the most precise estimate of the treatment effect.

Moderation by cohort and teacher? The treatment effect on African Americans' academic grades was consistent across different teachers and cohorts. No interaction effect involving experimental condition on the one hand and teacher team, cohort, or the combination of the two, was found for them [| *t*'s | < 1.19, all *P*'s >.30].

Repeated measures analyses of GPA-over-time data. We computed the linear trend by submitting GPA from the previous year (pre-intervention), average GPA in Year 1, and average GPA in Year 2 to a repeated measures analysis. The latter two average GPA indices consisted of the average GPA over each of the four terms of Year 1 and Year 2, respectively. The model testing the difference in linear trend as a function of condition controlled for cohort, teacher team, and student gender. The model testing the difference in linear trend as a function of both condition and students' racial group additionally included all 2-way interactions involving students' racial group, student gender, and experimental condition.

Effects involving pre-intervention performance levels. In our analysis of the effect of condition as a function of pre-intervention performance, we first standardized the two statistically significant pre-intervention performance covariates—previous year's GPA and pre-intervention 7th grade performance—and then averaged the two standardized scores. We then followed conventional regression and simple slopes analyses for testing interactions involving this variable (*S9, S10*). Specifically, pre-intervention performance was mean-centered on the sample being analyzed. That is, the sample mean was subtracted from each participant's score on pre-intervention performance. In testing the 3-way interaction involving condition, racial group, and pre-intervention performance, the latter was mean-centered using the sample mean for European Americans and African Americans combined. The main effect of pre-intervention performance was added to the basic regression model, replacing the two performance covariates on which it was based. All 2-way interactions involving racial group, condition, and pre-intervention performance were included in the model, as was the 3-way interaction.

In testing the 2-way interaction involving condition and pre-intervention performance among African Americans, pre-intervention performance was mean-centered using the sample mean for African Americans. To compute the experimental condition effect for low-performing African Americans, we added 0.675 standard deviations to the mean-centered pre-intervention performance measure of African Americans (*S9, S10*). Using this pre-intervention performance metric in the regression model on African Americans' data yields a condition main effect corresponding to its effect at the 25th percentile of pre-intervention performance (*S9, S10*). Likewise, to compute the experimental effect for high-performing African Americans, we subtracted 0.675 standard deviations from the mean-centered pre-intervention performance

measure among African Americans. Using this pre-intervention performance metric in the regression model on African Americans' data yields a condition main effect corresponding to its effect at the 75th percentile of pre-intervention performance. The resulting analyses test the condition effect among African Americans low and high in pre-intervention performance relative to their racial group.

Computation of the raw means in Figures 1 and 2 was accomplished by first performing a median split on pre-intervention performance separately for African Americans, so that the pre-intervention performance level (low vs. high) would represent students' performance relative to their racial group. The average GPA for each of these two performance subgroups thus reflects GPA at approximately the 25th and 75th percentile of pre-intervention performance. Mean GPA was then computed for African Americans' two pre-intervention performance groupings in each condition. Computation of adjusted means was accomplished by conducting separate analyses of covariance (ANCOVAs) for each of the two pre-intervention performance groupings among African Americans, using the same covariate variables and control variables enumerated in the regression model described previously.

Analysis of dosage manipulation. Two orthogonal contrast codes were created to test the dosage manipulation (*S9*). The first tested the effect of dosage, and assigned +1 to the low dose treatment condition, -1 to the high dose treatment condition, and 0 to the control condition. The second tested the effect of treatment, assigning -2 to the control condition and +1 to each of the two affirmation dosage conditions. Dropping the control condition from analyses and simply testing the difference between the two dosage conditions yields virtually identical results.

Likewise, an ANCOVA on Year 2 GPA, using experimental condition as a three-level factor (control, low dose treatment, high dose treatment), yielded a significant racial group x condition interaction and a significant effect of condition among African Americans. The follow-up tests among this racial group indicated a significant effect of receiving the affirmation treatment, but no significant effect of receiving low versus high dosage of it.

Data on remediation and advanced placement in math. Because of low counts, logistic and ordinal regressions on these outcomes were simplified to include only experimental condition and, where appropriate, students' racial group and the racial group x condition interaction, as well as significant pre-intervention performance covariates. Using the full models, where 7th grade team, gender, and cohort are retained, tended to yield stronger results on the remediation outcome, and did not change the results on the advanced math placement outcome (see Supporting Text for the latter results). The racial group x treatment interaction, and the treatment main effect among African Americans, were determined by assessing the change in model fit between the full model, that is with the critical effect, and the reduced model, that is without the critical effect, using the difference in the likelihood statistics (-2 log likelihoods).

The reported effect of students' racial group on remediation made use of a simple chi-square contingency table, and did not control for prior performance. Controlling for prior performance in testing an effect of racial group would be misleading, as it would attenuate the effect of racial group by removing from its effect variance attributable to racial differences in prior performance.

The condition effect for low-performing African Americans is 5% in the affirmation condition vs. 18% in the control condition [$\Delta\chi^2$ (1) = 8.98, P = 0.003].

Data from self-perceived adequacy measure. The analysis of the intervention's effect on African American students' self-perceived adequacy first involved determining students' level of pre-intervention performance relative to others in their racial group. This was accomplished by performing a median split based on students' 7th grade pre-intervention performance for each racial group, yielding groups of low and high performers in each racial group. The latter measure was used for generating these variables, as we were interested in the effect of poor performance early in the academic year on self-perceived adequacy at year's end. A paired t-test testing the difference between post-intervention (time 2) perceptions of adequacy and their pre-intervention (time 1) perceptions of adequacy was then conducted on low-performing African Americans, those benefiting most from the intervention, in each of the two experimental conditions.

Conventional regression and simple slopes analyses were used in the analysis of the relationship between 7th grade pre-intervention performance and subsequent adequacy (*S9, S10*). The outcome variable was post-intervention self-perceived adequacy. As in the analyses of performance, the regression controlled for wave, teacher team, and gender. The regression also controlled for baseline (pre-intervention) perceptions of adequacy. Pre-intervention performance was mean-centered on the sample of interest by subtracting the sample mean from each participant's score on pre-intervention performance.

In testing the 3-way interaction involving condition, racial group, and pre-intervention performance, reported in Endnote #18, the latter was mean-centered using the sample mean for European Americans and African Americans combined. The main effect of pre-intervention 7th grade performance was included in the model. All 2-way interactions involving racial group, condition, and pre-intervention performance were also included in the model, as was the 3-way interaction.

In testing the 2-way interaction involving condition and pre-intervention performance among African Americans, reported in Endnote #18, pre-intervention performance was mean-centered using the sample mean for African Americans. To compute the effect of pre-intervention performance for African Americans in the control condition, we used the original dummy-coded condition variable, in which the control condition was designated as 0, the affirmation condition as +1. To compute the effect of pre-intervention performance for African Americans in the affirmation condition, we reversed the values of this variable, such that the control condition was assigned a value of +1 and the affirmation condition assigned a value of 0. Using these values in the regression model on African Americans' data yields the coefficient associated with pre-intervention performance in the control condition and affirmation condition, respectively (*S9*). To compute the experimental condition effect for low-performing African Americans, we added 0.675 standard deviations to the mean-centered pre-intervention performance measure of African Americans (*S9, S10*). Using this pre-intervention performance metric in the regression model on African Americans' data yields a condition main effect corresponding to its effect at the 25th percentile of pre-intervention performance (*S9, S10*).

Eight students provided post-intervention adequacy data but no pre-intervention adequacy data. To avoid losing their data, each missing pre-intervention value was replaced with the sample mean, and a dummy variable to code whether the value was missing or not was also included in the model. As the other baseline variables proved only modestly correlated with the relevant predictor, the highest correlation being $R = 0.33$, $P < 0.001$, we opted to use this strategy rather than using regression-based estimates derived from the other baseline variables for the missing values. In fact, results were unaffected when regression-estimated values were used for these missing values, or when participants with missing pre-intervention values were simply dropped from analyses.

Supporting Text

Additional Analyses

Robustness of basic analytic effects when interactions involving pre-intervention performance are included in the model. In analyses of average two-year GPA, the overall racial group x treatment interaction retains significance of virtually identical magnitude when the 3-way interaction with pre-intervention performance is entered in the model. The racial group x condition interaction corresponded to 0.35 grade points, $t(319) = 3.21$, $P = 0.001$. Additionally, the treatment main effect among African Americans corresponded to 0.28 grade points, $t(144) = 3.79$, $P < .001$.

In these analyses, all continuous predictors are mean-centered on the sample being analyzed. All binary variables are contrast-coded so as to average to 0, including experimental condition (control = -1; affirmation = +1) and student group (European Americans = -1; African Americans = +1). Doing so preserves the interpretability of the lower-order main effects and interactions (*S9, S10*). These results indicate that the affirmation effect is sufficiently robust to be apparent at average pre-intervention performance levels.

Mediation of Year 2 treatment effects by Year 1 treatment effects. Conventional mediational testing (*S11*) yielded evidence that the treatment's effect on Year 2 GPA was partially mediated by its effect on Year 1 GPA. When entered into the original regression model predicting Year 2 GPA, Year 1 GPA was highly predictive of Year 2 GPA for all students, even with pre-

intervention performance controlled [$B = 0.64$, $t(320) = 10.46$, $P < 0.001$]. Additionally, the significant racial group x treatment interaction on Year 2 GPA fell from $B = 0.39$, $P = 0.001$, to $B = 0.21$, $t(320) = 2.01$, $P = 0.046$, when Year 1 GPA was controlled. This represents a 46% reduction. Likewise, the treatment effect among African Americans fell from highly significant [$B = 0.27$, $P = 0.003$] to marginal [$B = 0.15$, $t(143) = 1.88$, $P = 0.063$], a 44% reduction. Sobel tests confirmed that the reductions in the interaction and condition effects were both significant [*Z's* = 2.84 and 2.60, *P*'s = 0.005 and 0.009, respectively]. This mediational analysis supports the notion that the intervention's short-term effects on performance mediated its longer-term effects.

Of course, this mediational analysis is, like conventional mediational testing procedures in psychology, fundamentally correlational. Thus alternative accounts could explain the results. For instance, extraneous third variables correlated with the error terms of both the mediator (Year 1 GPA) and the outcome (Year 2 GPA) could be in play. There may be a number of reasons for why those who initially benefit from an intervention subsequently display benefits. As noted by recent reports, in the absence of experimental manipulation of the candidate mediator, such limitations are inherent in conventional, regression-based mediational testing procedures (*S12, S13*), as are others that may obscure mediation, such as collinearity and suppression effects. In order to, at least partially, address the issue of potential third variables being in play, we controlled for various baseline measures, including baseline performance, which may affect responsiveness to the intervention. Holding these preexisting individual and environmental factors constant, the analysis indicates that performing relatively better (or worse) in Year 1 predicts performing better (or worse) in Year 2, with the former mediating the intervention's effect on the latter. Given some of the drawbacks of non-experimental approaches to mediation,

these results regarding the mediational role of short-term performance benefits are suggestive but not conclusive.

On the other hand, these results provide relatively strong evidence of the recursive processes that we suggest underlie the intervention's long-term effects when viewed in conjunction with the results involving the treatment booster manipulation, the moderating role of prior performance, the treatment's effect on the downward performance trend over time and its weakening of the relationship between early poor performance and subsequent outcomes, and the effects on self-perceived adequacy. The mediating role of short-term performance in long-term performance is also grounded in a large body of previous research findings that are consistent with the posited mechanism (*S14*). This research indicates that performance affects performance—that failure, especially when persistent and construed negatively, undermines people's self-efficacy, motivation, and performance, and that success often enhances them. Future research that directly manipulates mediational variables could shed further light on these processes (*S12, S13*).

Mediational role of self-perceived adequacy? While higher self-perceived adequacy correlated with higher GPA, evidence that it statistically mediated the treatment effect on performance was not consistently found. Future research could address this issue more definitively through more extensive in-class measurement of mediational states and through targeted experimental manipulation of the candidate mediator (*S12, S13*). It should be noted that a lack of mediation would be informative, as it would suggest that the affirmation may have separable effects on a variety of educationally relevant psychological and behavioral outcomes.

Effects on advanced placement in math. Students could take an advanced math class in both 7th and 8th grades. This was the only core course with basic and advanced curricula. Students could shift levels in the transition between 7th and 8th grade. Among African Americans in the control condition, none raised a level while 6% dropped a level. In contrast, among affirmation-treated African Americans, 4% raised a level while only 2% dropped. For European Americans, there was no variation by condition, with 6% dropping a level, 1% raising a level. The latter may have occurred because a larger number of European Americans than African Americans began 7th grade in an advanced math class. Using whether students dropped, remained constant, or gained a level in math as an outcome, ordinal regression identified the familiar treatment x racial group interaction [$\Delta\chi^2$ (1) = 5.05, P = 0.025] and affirmation effect among African Americans [$\Delta\chi^2$ (1) = 3.93, P = 0.047], with no such effect for European Americans [$\Delta\chi^2$ (1) = 0.94, P = 0.332].

Did the treatment booster benefit the Year 2 GPA of affirmation-treated African Americans who displayed relatively little treatment benefit in their Year 1 GPA? The answer is no. A multiple regression found that the null effect of the treatment booster among African Americans was consistent across level of performance in Year 1. That is, when Year 1 GPA and its interaction with condition were added to the basic regression model predicting African Americans' Year 2 GPA, there was no interaction between the treatment booster contrast code and Year 1 GPA. This was the case regardless of whether preintervention performance was included in the model or not [| t's | s < 1.3, P's > 0.22].

Use of hierarchical linear modeling (HLM) to determine treatment effect. Our regression analyses attended to classroom and cohort effects by using dummy variables. Because

randomization occurred at the level of individual student, with each student in each classroom randomly assigned to experimental condition, the multiple regression approach is unbiased in its estimate of condition effects. It is also possible to attend to classroom level effects through the use of HLM to take into account the nesting of students within teacher. Using HLM 6.03, we tested a random-intercept model where student-level variables were nested within a combination of teacher and wave (*S15*). Because randomization was at the level of individual student, there should be little difference in the results of HLM and multiple regression. This was the case. The HLM-estimated parameters for the racial group X treatment interaction and the treatment effect for African Americans on overall GPA were very similar to those found with multiple regression [racial group x condition interaction: $\gamma = 0.31$, $t(329) = 3.32$, $P = 0.001$; condition effect for African Americans: $\gamma = 0.23$, $t(152) = 3.39$, $P = 0.001$].

Supporting Figure

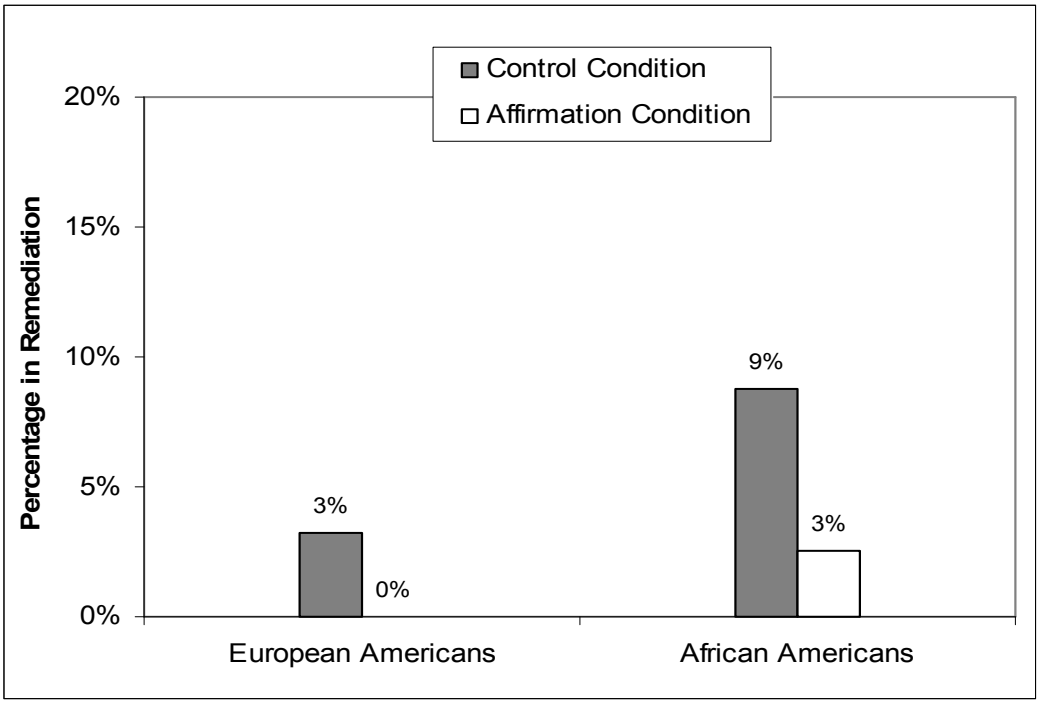

Fig. S1. Percentage of students placed in remediation by school (assigned to a remedial program or held back in their grade), as a function of student group (African American vs. European American) and experimental condition.

Supporting Table

Table S1. Summary of Regression Model Predicting Average Two-Year GPA

Variable	*B*	*SE*	*t*
Student Racial Group	-0.08	.09	-0.94
Student Gender	0.16	.08	2.08*
Cohort Code 1	-0.10	.06	-1.81~
Cohort Code 2	0.04	.06	0.79
Year 1 Teacher Code 1	0.15	.10	1.63
Year 1 Teacher Code 2	0.11	.10	1.19
Year 2 Teacher Code 1	-0.04	.43	-0.10
Year 2 Teacher Code 2	-0.12	.43	-0.27
Year 2 Teacher Code 3	-0.09	.44	-0.21
Year 2 Teacher Code 4	0.71	.48	1.50
Pre-Intervention Seventh Grade Performance (Standardized)	0.22	.03	6.72**
GPA from Previous Year	0.88	.04	21.04**
Experimental Condition	-0.02	.07	-0.22
Racial Group x Gender	-0.14	.09	-1.51
Gender x Condition	-0.13	.09	-1.39
Racial Group x Condition	0.33	.09	3.59**

Note: Student racial group code: 0 = European Americans, 1 = African Americans; student gender code: 0 = male, 1 = female; teacher code variables were dummy-coded (0/1) variables used to designate participating teams of teachers.

~ $P = .07$ * $P < 0.05$ ** $P < 0.01$

Supporting References

S1. G. L. Cohen, J. Garcia, N. Apfel, A. Master. *Science,* **313**, 1307 (2006).

S2. J. Aronson, in *Improving Academic Achievement: Impact of Psychological Factors on Education* (Academic Press, San Diego, CA, 2002), pp. 303-S3. S. Fein, S. J. Spencer, *J. Pers. Soc. Psychol.*, **73**, 31 (1997).

S3. S. Fein, S. J. Spencer, *J. Pers. Soc. Psychol.,* **73,** 31, 1997.

S4. G. L. Cohen, J. Aronson, C. M. Steele, *Pers. Soc. Psychol. Bull.,* **26**, 1151 (2000).

S5. C. M. Steele, in *Advances in Experimental Social Psychology,* L. Berkowitz, Ed. (Academic Press, New York, 1988), pp. 261-302.

S6. D. K. Sherman, G. L. Cohen, in *Advances in Experimental Social Psychology,* M. P. Zanna, Ed. (Academic Press, San Diego, CA, 2006), pp. 183-242.

S7. G. R. Salancik, M. Conway, *J. Pers. Soc. Psychol.*, **32**, 829 (1975).

S8. G. M. Walton, G. L. Cohen. *J. Pers. Soc. Psychol.,* **92**, 82 (2007).

S9. C. M. Judd, G. H. McClelland, *Data Analysis: A Model Comparison Approach* (Harcourt, Brace, Jovanovich, San Diego, 1989).

S10. J. Cohen, P. Cohen., L. S. Aiken, S. G. West. *Applied Multiple Regression – Correlation Analysis for the Behavioral Sciences* (Lawrence Erlbaum, Mahwah, NJ, 2002).

S11. R. M. Baron, D. A. Kenny. *J. Pers. Soc. Psychol.,* **51**, 1173 (1986).

S12. J. G. Bullock, D. P. Green, S. E. Ha (http://www.ipeg.org.uk/events/field_experiments/Documents/A%20Critique%20of%20Conventional%20Mediation%20Analyses%20--%20Bullock%20Green%20and%20Ha.pdf)

S13. S. J. Spencer, M. P. Zanna, G. T. Fong. *J. Pers. Soc. Psychol.,* **89**, 845 (2005).

S14. A. Bandura. *Self-Efficacy: The Exercise of Control* (W. H. Freeman and Company, New York, 1997).

S15. S. W. Raudenbush, A. S. Bryk, Y. F. Cheong, R. T. Congdon. *HLM 6: Hierarchical Linear and Nonlinear Modeling* (Lincolnwood, IL: Scientific Software International, 2004).