Interaction, Stereotypes and Performance. Evidence from South Africa

Lucia Corno, Eliana La Ferrara, Justine Burns*

This version: March 2022

Abstract

We exploit a policy designed to randomly allocate roommates in a large South African university to investigate whether inter-racial interaction affects stereotypes, attitudes and performance. Using Implicit Association Tests, we find that living with a roommate of a different race reduces white students' negative stereotypes towards black ones and increases inter-racial friendships. Interaction also affects academic outcomes: black students improve their GPA, pass more exams and have lower dropout rates. This effect is not driven by roommate's ability. Students who benefited the most from inter-group contact in terms of performance were relatively poorer and had roommates with less negative stereotypes.

*We thank Michela Carlana, Nicola Gennaioli, Caroline Hoxby, Michael Kremer, Monica Martinez-Bravo, Rohini Pande, Enrico Rettore, Piero Veronese and seminar participants at NBER Political Economy, CIFAR at LSE, IEA World Congress, CEPR Public Policy, EUDN, Boston University, Tufts, University of Chicago, University of Turin, SciencesPo, INSEAD, Oxford, Universidade Nova de Lisboa, Pompeu Fabra, EUI, University of Essex, IRVAPP, Stockholm University, University of Cape Town, CERGE-EI and Hebrew University for helpful comments. We are grateful to Selene Ghisolfi for invaluable help with data collection and analysis. Erick Baumgartner, Nicola Fontana, Jack Melbourne, Vrinda Kapoor and Martina Zanella provided excellent research assistance. La Ferrara gratefully acknowledges financial support from the ERC Advanced Grant ASNODEV (Contract No. 694882). The usual disclaimer applies. Corno: Cattolica University and LEAP, lucia.corno@unicatt.it; La Ferrara: Bocconi University, LEAP and IGIER, eliana.laferrara@unibocconi.it; Burns: University of Cape Town, justine.burns@uct.ac.za

1 Introduction

We study whether a policy designed to exogenously generate exposure to members of different racial groups induces changes in attitudes and stereotypes, and whether this translates into sizeable performance gains for the individuals initially subject to negative stereotypes. We study this in the context of South Africa, a country where the experience of apartheid made people relatively prone to stereotyping and led to the economic marginalization of black South Africans. We take advantage of a policy implemented by the University of Cape Town (UCT) with the aim of promoting racial integration. This policy randomly allocates students across university residences and –in some of the residences– to roommates, thus providing a unique opportunity to estimate the causal effect of a roommate's race on individual attitudes and behavior.

We address two sets of questions: (i) Does interaction with someone of a different race change individual stereotypes towards that race? (ii) What are the effects of inter-group contact on academic outcomes? Do stereotypes play a role in explaining academic effects? To the best of our knowledge, ours is the first study to contextually estimate the effect of inter-group contact on stereotypes and performance.

Whether increased interaction with members of other groups would increase or decrease stereotypes is theoretically unclear. Allport's (1954) 'contact hypothesis' maintained that, under certain conditions, exposure to outgroup members would allow individuals to better understand their characteristics and points of view, thus diminishing reliance on stereotypes and eventually ameliorating inter-group relations.[1] On the other hand, a recent literature on the 'negative contact hypothesis' contends that negative contact makes categories more salient than positive contact, potentially leading to an increase –as opposed to a reduction– in negative outgroup stereotyping (Paolini et al., 2010; Barlow et al., 2012).[2]

To address the above questions, we collected data on students living in UCT residences who were exposed to the policy of random roommate allocation. Our sample includes 499 freshmen living in double rooms, whom we interviewed at the beginning and at the end of the 2012 academic year. Our first outcome of interest is prejudice or stereotype held against members of different racial groups. To gain a possibly ob-

[1] The conditions that need to be satisfied in Allport's view include the fact that groups enjoy equal status in the relationship, share common goals, benefit from cooperation and that interaction is supported by a commonly recognized authority. All of these conditions can be said to apply in the university setting we study.

[2] For a theoretical model of stereotype formation, see Bordalo et al. (2016).

jective measure, we administered a series of Implicit Association Tests (IATs). The IAT was first introduced by Greenwald and Banaji (1995) and is a tool used by social psychologists. It exploits variation in the time that individuals take to complete a rapid categorization task that involves associating concepts with visual cues about race. The underlying idea is that subjects who are systematically slower in associating certain pairs implicitly reveal mental processes that tend to perceive those pairs as less common. In addition to the standard "Race IAT", which elicits associations between generally "positive" concepts and race, we designed an IAT to elicit associations between academic ability and race –we refer to this as the "Academic IAT". The advantage of IATs over self-reported measures of prejudice is significant, especially in contexts where subjects may be reluctant to disclose prejudice or may not be fully aware of it. Although criticisms exist within the psychology literature on the predictive power and reliability of IAT scores (see Section 2.3), the IAT remains widely used in social psychology as a measure of implicit bias. The IAT has also been used in the economic literature, although –to the best of our knowledge– we are the first to use it as an outcome to estimate the impact of a policy to promote integration on racial stereotypes.[3]

Our first main result shows that exposure to members of a different race led to significant changes in stereotypes: white students became relatively less prejudiced against black students. The effect is sizeable, corresponding to .63 of a standard deviation of the *Race IAT*. The magnitude of the estimated effect suggests that the treatment would close the gap in stereotypes between white students and black students, as measured by the *Race IAT*. We do not find corresponding changes in the *Academic IAT*, suggesting that interaction *per se* does not necessarily induce an update in beliefs on ability in a direction that is favorable to black South Africans.

Our second set of results concerns academic performance. We find significant effects of inter-group contact on academic achievement, heterogeneous across groups. Grade point average (GPA) scores improve by .26 standard deviations for black students sharing a room with non-black students. This closes over 1/4 of the gap in GPA between white and black respondents. Black students in mixed rooms also pass a higher number of exams and are more likely to be eligible to continue to the following year. The

[3]Beaman et al. (2009) use an activity-based IAT to estimate the effect of exposure to female leaders on gender stereotypes in India. Lowes et al. (2015) use the IAT to provide descriptive evidence on implicit attitudes towards different ethnic groups in the Democratic Republic of Congo. Barnhardt (2009) exploits a natural experiment with public housing in India to study the effects of geographic proximity on religious prejudice using an IAT for the categories of Hindu and Muslim.

positive effect on academic performance is long lasting: it persists in the second year, when most of the students are no longer in residences. No significant impact on academic performance is found for white students in mixed rooms: the estimated effects are virtually zero and statistically insignificant.[4]

Interestingly, the positive effect on black students' performance is not driven by 'standard' academic peer effects. While it is true that, due to affirmative action at UCT, a black exposed to a white is on average exposed to someone with a higher entrance score, this does not explain the performance effects. In all regressions we control for the roommate's admission score into UCT as a proxy for ability, and the latter is typically insignificant. Also, being in the same faculty or taking the same courses does not explain the positive performance effects for black students in mixed rooms.

To shed more light on the factors underlying the academic performance gains, we predict the heterogeneity of treatment effects using the machine learning tools proposed by Chernozhukov et al. (2020). When we compare the individuals who benefit relatively more from being in mixed rooms (top 20 percent of the predicted Conditional Average Treatment Effects) to those in the bottom quintile of CATE, we find that the former are significantly poorer, less likely to have attended a private high school, and have roommates with lower admission score and less negative stereotypes (as measured by their baseline IATs).

The latter result is particularly interesting, as the lower prejudice of the roommate may have helped reduce inter-group anxiety (Stephan and Stephan, 1985), thus allowing students to perform better academically. While previous work has examined the effect of random roommate allocation on academic performance, to our knowledge ours is the first attempt to study how this effect varies with the stereotypes held by the subjects involved. If academic gains from interaction are to be realized, it seems plausible that stereotypes would play a significant role in making the interaction viable and beneficial for the parties involved.

To further explore the mechanisms of interaction, in the last part of the paper we examine the effects of a roommate's race on a variety of attitudinal and behavioral measures. We find that exposure to a roommate of a different race increases interracial interactions outside the room: students in mixed rooms report that they hang

[4]We estimate an (insignificant) 0.03 standard deviation reduction in GPA, an (insignificant) 0.07 standard deviation reduction in the number of exams passed, and an (insignificant) 5 percentage point increase in the probability of being eligible to continue, starting from a baseline of 0.92.

out more often with people of a different race, and their desired and actual number of friends and study-mates of a different race is higher than that of students in same-race rooms. Self-reported attitudes also improve, as students in mixed rooms report talking more frequently about race and feeling more comfortable about it. They also feel less self-conscious dancing with or dating a person of a different race. We find weaker evidence that students paired with roommates from other groups exhibit more generalized prosocial behavior, as measured by participation in voluntary organizations and social services, and cooperative behavior in a prisoner's dilemma game.

Research in economics has widely studied the effects of ethnic diversity and racial segregation. In this literature, integration policies are often proposed as a means of reducing racial gaps in outcomes, reducing the costs of ethnic divisions and leveraging the potential benefits of diversity.[5] Social psychologists, on the other hand, have studied diversity and underlined the importance of identity and stereotype formation. In this literature, the main role served by integration is not that of improving (economic) performance, but of changing individual attitudes and stereotypes, possibly reducing prejudice and inter-group conflict.[6] Our paper is an attempt to bring together these two sides of the problem to show that attitudinal change and performance gains may go hand-in-hand.

Our work relates to three strands of literature. The first is the literature on the effects of integration policies on inter-group attitudes. Boisjoly et al. (2006) find that randomly matching roommates of different races in a US university increases support for affirmative action and empathy towards other groups. Van Laar et al. (2005) use housing assignments of first-year college students at University of California and find that having a roommate from another ethnic group decreases prejudice. Carrell et al. (2019) exploit data from the United States Air Force Academy to show that white males randomly assigned to both a higher number, as well as higher-aptitude, black peers in their freshman year are more likely to match with a black roommate in their sophomore year. Finally, inter-group contact of different religious groups, in particular Christian and Muslim (e.g., Scacco and Warren, 2018; Mousa, 2020) and of people from different

[5]For a review of the literature on ethnic diversity, see Alesina and La Ferrara (2005). A recent study of the economic costs of ethnic divisions is Hjort (2014), while a challenge to the assumption of coethnic bias is presented by Berge et al. (2015). On diversity-enhancing policies, see, among others, Fryer and Loury (2013) for a theoretical analysis and Kling et al. (2007) for an evaluation of the Moving to Opportunity program.

[6]For a meta-analysis of the relation between inter-group contact and prejudice, see Pettigrew and Tropp (2006) and Paluck, Green and Green (2018).

castes (Lowe, 2021) increases positive attitudes and inter-group friendship.

The above papers use self-reported attitudes and, some of them, behavioral outcomes. We complement explicit attitudes and behavioral outcomes with Implicit Association Tests. Glover, Pallais and Pariente (2017) examine the performance of cashiers in a French grocery store chain and find that minorities perform more poorly when they work with managers biased against them (as determined by IATs). Our setting is different from the latter, as it is one where interaction occurs among peers, as opposed to people at different levels of the hierarchy, and according to Allport's (1954) hypothesis this should be key for realizing the gains from contact. Also, we look at a different set of outcomes than Glover et al. (2017), including stereotypes, social interaction and academic performance.

Recent papers by Barnhardt (2009) and Rao (2019) study different forms of integration in India. Barnhardt (2009) examines the effects of neighborhood religious composition on inter-religious attitudes, while Rao (2019) studies the impact of changes in the wealth composition of children's classmates on prosocial behavior and test scores in Delhi's private schools. With respect to these authors, we focus on a different dimension (i.e., race as opposed to religion or social class) and we also simultaneously study prejudice reduction and educational outcomes.

A second body of literature brings the notion of identity to the forefront of economic analysis, embedding concepts developed in the social psychology literature, including the seminal work by Tajfel et al. (1971). Contributions include, among others, Akerlof and Kranton (2000), Hoff and Pandey (2006) and Shayo (2009). While we do not directly elicit notions of self-identification, some of our results on the salience of race and on revised beliefs regarding out-group members speak to the issue of identity formation.

Finally, our paper contributes to the vast literature on the effect of peers on human capital. Most of this literature studies the effect of peers' ability and academic performance (Sacerdote, 2001; Lyle, 2009; Hanushek, Kain and Rivkin, 2009; Garlick, 2018; Hasan and Badge, 2013) and derives implications for policies such as tracking (Duflo et al., 2011; Carrell, Sacerdote and West, 2013). While we do not estimate 'endogenous' peer effects, our paper aims at shedding light on the role played by racial stereotypes in academic interactions.

The remainder of the paper is organized as follows. Section 2 provides some institutional background and describes the data we collected. Section 3 shows descriptive statistics and discusses the identifying assumption underlying our work. In section 4 we present our empirical strategy. Section 5 contains the econometric results and section

6 concludes.

2 Background and data collection

2.1 Institutional setting

The University of Cape Town (UCT) is a public research university located in Cape Town. It is the oldest and most prestigious university in South Africa and it enrols approximately 5000 incoming freshmen every year, more than half of whom live in university residences. Incoming students were historically allocated into dormitories to live with students whose academic performance in standardized high school graduation tests was similar to their own. This tracking regime was replaced in 2006 with a policy of randomly assigning incoming students to dormitories.[7]

Students submit applications to the university between July and October to start studying in January of the following year. UCT's admission policy is mainly based on a measure called Admission Points Score, computed from the high school grades in the final school year, but it is also designed in order to build a student body that reflects the demographics of South African society. In 2012 the composition of the incoming freshmen class was: 36 percent black, 16 percent coloured, 40 percent white, and 8 percent Indian, Asian or other race.[8]

In the application form, students may request to live in university residences. Only students living outside the Cape Town area can apply for accommodation. Exceptions are made for disadvantaged students or for those with great academic merit. The policy and criteria for admission to UCT student housing assume that a first year student will enter a first-tier (catering) residence and in subsequent years move to second- or third-tier (senior catering or self-catering) accommodation. While second-year students may express a preference for the residence they wish to be assigned to, freshmen assignment to residences relies on a random allocation system.

Once first-year students are randomly allocated to one of the fifteen residences, they are assigned a room which can be either single or double occupancy. All rooms are single gender. The criteria for allocation to specific rooms within the residence are decided by the warden and may vary by residence.[9] Our analysis focuses exclusively on

[7] Garlick (2018) compares the two policies in terms of GPA distributions using a difference-in-differences strategy.

[8] In South Africa, the term 'coloured' denotes people of mixed ethnic origin who may have diverse ancestries, including European, Asian and local Khosa and Bantu ethnic groups.

[9] Therefore the random allocation of students across dormitories is a policy at the university level,

eight residences with double rooms in which during the year of our study the wardens stated that they implemented a random allocation mechanism, conditional on gender, and where white and black students coexist.[10]

Approximately one week before the beginning of the academic year, each residence organizes an open day with first year students to explain the rules and features of the residence. During the open day, each student is assigned to a room. In some residences, the random assignment takes the form of extracting the room number from an urn: if the room extracted is a single room, the number is removed from the urn; if it is a double room, it is placed back in the urn so that a roommate may extract it again. In other residences the wardens randomly select students and their roommates from the list of students enrolled in the residence.[11]

Approximately 50 percent of undergraduate students living on campus are in shared rooms in their first year in residence. The Residence Management Services is in charge of residence applications; it records, for each student, his/her room number and the dates in which he/she moved in or out of the residence. Rooms are never reserved irrevocably and may be switched. First year students may also decide to swap residences. In our sample 19 percent of the students interviewed at follow-up declare that they changed roommate since the beginning of the year. In all our analysis we will use the initial room assignment, thus reporting "intention to treat" estimates.[12]

2.2 Sample

Our working sample includes one cohort of students: 499 freshmen who joined UCT in 2012 and who were assigned to double rooms in any of the eight residences that reported randomly allocating rooms. We conducted two rounds of data collection: a baseline in February 2012, during the first week of the academic year, and a follow-up survey in September 2012, near the end of the academic year and just before students took

while the random allocation of roommates is a discretionary policy at the dormitory level.

[10] We originally collected data for ten residences that told us they were implementing a random room allocation scheme, but two of the residences had a very small number of double rooms and did not have any white student in double rooms. We drop these residences, which accounted for a total of 9 observations in the original sample. Our results are unchanged if we keep them in the analysis.

[11] It is possible that wardens may 'adjust' the composition of some rooms, e.g., to ensure that each floor or wing has a certain composition. While this was not described to us as a standard procedure, we cannot rule it out. We will however provide evidence that such exceptions, if they occurred, did not lead to significant deviations from a random allocation in terms of most observable characteristics.

[12] Note that, although 19 percent of the students change room or roommate, only 2 students go from a mixed race room to a non-mixed room, and no one goes from non-mixed to mixed. For this reason, we do not report local average treatment effect (LATE) estimates.

their final exams.[13] As part of the data collection, we conducted a series of Implicit Association Tests (IATs) both at baseline and at follow up. During the follow up survey, besides collecting data through questionnaires and IATs, we also conducted lab experiments with the same individuals who took the baseline survey.[14]

Students were recruited to participate in the project through a variety of channels. First, the project was advertised during a residence meeting among wardens and students. Second, posters advertising the project were hung in visible places (e.g., residence halls) about one week before the kick-off. Third, we sent an e-mail to all the students in the participating residences to schedule an appointment for the survey at their most convenient time. In all cases we described the goal of our research as being about "student life at UCT", without mentioning race.

The survey questionnaire, the IATs and the experimental game were conducted in each residence on laptops and under the supervision of two enumerators per residence. We did our best to ensure no communication among students during the survey. To try not to contaminate the IATs scores, we conducted them first, followed by the survey questionnaire and then by the experimental game. For their participation in the survey, each student received a monetary incentive of 30 South African Rands, i.e., approximately 3.5 US dollars (corresponding to 1/2 hour of an enumerator's wage).

Our initial sample of survey respondents for whom we had non-missing IAT scores at baseline was 621 freshmen.[15] Of these, 499 were successfully traced at follow-up (with a tracking rate of 80 percent). In Appendix Table A1 we examine the correlates of the decision to participate in the follow-up round. We find no differential attrition between respondents in mixed and non mixed rooms (column 1). Importantly, attrition does not depend on our measures of stereotypes, the *Race IAT* (column 2) and the *Academic IAT* score (column 3), nor on the interaction between these scores and *Mixed Room* (columns 4 and 5) and between IAT score and race dummies (columns 6 and 7). Looking at other controls, it emerges that women and richer students (as measured by

[13] Students typically move into the residence in the week before the start of the academic term (orientation week). This implies that in general they would have been exposed to their roommate for a minimum of 0 and a maximum of 10 days before they took the baseline survey, depending on the day they moved in and on when exactly the data collection was done in their residence.

[14] The order of the sections was: (IAT, survey) at baseline, and (IAT, survey, experimental game) at follow-up. Note that even in the presence of order effects, these should not systematically differ between people assigned to mixed versus same race rooms.

[15] The sample size of 499 students mentioned above refers to our final working sample, i.e., students for whom we have both baseline and follow-up survey data with non-missing values for the *Academic IAT* and *Race IAT*.

their consumption) are relatively less likely to participate in the follow-up survey. The latter result may be due to the relatively low monetary incentives for participation.

Appendix Table A2 compares our sample to other freshmen living in the same residences but who did not take the survey, as well as to all freshmen (including those not in residences) and the total student population at UCT. While the students in our sample are broadly representative of freshmen living in residences, they differ from students living off campus and from other years in that they have a higher UCT admission score and they are more likely to be female. This is consistent with the fact that merit is one of the admission criteria in UCT residences.

2.3 Implicit association tests

Our key outcome of interest is racial stereotypes, as measured by the Implicit Association Test (IAT). The IAT is an experimental method introduced by Greenwald and Banaji (1995) and Greenwald, McGhee and Schwartz (1998), based on the idea that respondents who more rapidly pair two concepts in a rapid categorization task more strongly associate those concepts (e.g., how fast people pair images of black versus white people with descriptions of leadership roles). Slower speed in associating certain pairs denotes mental processes that tend to perceive those pairs as less common. This tool has been widely employed in social psychology to understand implicit cognition, that is, cognitive processes of which an individual may not be aware and that include, among others, perception, stereotyping, and memory. For our purposes, IATs have the advantage of avoiding social desirability bias in self-reported answers and implicitly reveal attitudes that individuals may be uncomfortable disclosing or not fully aware of, such as racial prejudice. This is particularly relevant in the South African context. Thus, we use IATs to complement subjective and self-reported perceptions of inter-ethnic attitudes with more 'objective' measures of racial stereotypes.

We conducted two types of IATs. The first was a standard test in which tasks involved pairing positive and negative attributes (e.g., "happy", "good", "terrible", "failure") with the racial categories of White South African and Black South African. Following the literature (Blanton et al. 2009 and Oswald et al. 2013), we refer to this taste-based IAT as the '*Race IAT*'. The second IAT was less standard and we designed it to elicit associations between academic ability and race. We asked people to match pictures of different race with different percentiles of the grade distribution, corresponding to high and low performance in the UCT grading system. We denote this as the '*Academic IAT*'. The *Academic IAT* has exactly the same structure of the

Race IAT, but instead of categories of good vs. bad, the categories are 'First class (>75%)' vs. 'Third class (<60%)', and instead of attributes like 'happy' or 'terrible', we have percentiles like '78%' or '56%'. The goal of conducting this second IAT was to test whether differential interaction with members of the opposite race may reflect priors on how much one can benefit in terms of learning and academic success, based on the beliefs that one holds about the academic performance of the other race. In this sense, the *Academic IAT* is performance-based, rather than taste-based, and should be seen as a measure of beliefs regarding academic performance rather than a measure of prejudice. Online Appendix A details the procedures of the two IATs.

Figure 1 shows the density of the *Race IAT* score (panel A) and the *Academic IAT* score (panel B) at baseline, separately for white and black respondents.[16] We code the IAT so that lower values of the score indicate more negative stereotypes towards black relative to white individuals.

Two interesting patterns emerge from this figure. First, on average, all groups hold negative stereotypes towards black individuals, as the mean of the *Race* and the *Academic IAT* is always negative. Second, while white students are more prejudiced against black individuals when we consider the *Race IAT* (panel A), no significant difference across races emerges when looking at distributions of the *Academic IAT* (panel B). This is interesting because it suggests that the *Academic IAT* is more likely to reflect statistical information on the relative academic performance of the two groups, while the *Race IAT* may embed more elements of 'taste'.

IAT validity: discussion

While extensively used in the literature, the IAT is not uncontroversial and a vibrant debate exists among social psychologists about its interpretation and potential limitations. We cannot do justice to the complexity of the arguments here, but we discuss some key critiques of the IAT and the extent to which they may apply in our setting.

The first thing to note is that the IAT is an outcome in our analysis, so most of the existing critiques do not apply to our setting, because if the IAT it picked up pure noise it would be difficult for us to detect an effect, which we *do* find.

Among the critics, some have questioned the relevance of implicit attitudes for explaining discriminatory behavior (see Blanton et al., 2009 and Oswald et al., 2013 for a meta-analysis). However, the majority of the studies included in these meta-analyses

[16] The pattern when we pool black, coloured and other race in a single category is very similar to that of black students alone.

involve experiments in the lab and with fewer than 50 subjects.[17] On the other hand, a large number of papers have shown that implicit racial preferences predict actual behavior (see Greenwald et al., 2009 for a meta-analysis), including negative inter-racial contact (McConnell and Leibold, 2001), biases in medical decision making (Green et al., 2007), and hiring discrimination (Rooth, 2010). Beyond the social psychology literature, an increasing number of economists have used IATs to capture prejudice and stereotypes and found that it predicts behavior (e.g., Bertrand et al., 2005; Beaman et al., 2009; Reuben, Sapienza and Zingales, 2014; Alesina et al., 2018; Carlana, 2019; Avitzour et al., 2020).

In our data we find mixed evidence on the correlation between the IAT and behaviors or explicit attitudes. Appendix Table A3 (Panel A) shows that, among white students, higher values of the *Race IAT* (i.e., less negative stereotypes towards black students) are significantly and positively associated with the expectation that the partner in a prisoner game will cooperate. This holds also when conditioning on stated racial preferences (column 6), suggesting that the IAT has predictive power beyond self-reported preferences. The correlation is positive but not significant when we look at the probability of cooperating in the prisoner's dilemma.[18] When we look at explicit attitudes and reported inter-racial friendships (panel B), the results are weak, with the exception of a significant positive correlation with attitudes in the full sample. This is not surprising in a context where race is a sensitive issue and explicit measures may suffer from reporting bias.

A second critique of the IAT is that it may overestimate bias for individuals who are more consistent with their answers, since differences in reaction times are normalized by standard deviations in the scoring algorithm. This critique is not very relevant in our setting because, when we estimate the impact of being in a mixed room on stereotypes, we condition on the respondent's initial IAT score, thus capturing a change over time for the same respondent. There is no clear reason why people in mixed rooms should become less consistent in their answers over time.

Third, implicit measures of prejudice may reflect familiarity with negative, yet egalitarian, associations, such as the fact that certain racial groups have been oppressed or

[17] Also, Greenwald, Banaji and Nosek (2015) suggest that Oswald et al. (2013) estimate aggregate correlational effect sizes that are large enough to justify concluding that IAT measures predict societally important discrimination.

[18] The coefficient for black respondents goes in the opposite direction, as one would expect given the way in which the IAT score is oriented, but it is smaller and insignificant. The fact that the predictive power of the IAT is heterogeneous across groups is consistent with Kurdi et al. (2018).

victimized (Uhlmann, Brescoll and Paluck, 2006). In our case this would suggest that white students exposed to black roommates would appear more biased, while we find the opposite.

Fourth, the results of the IAT may not be stable over time, reflecting exposure to contextual short-term factors.[19] Again, this is not a serious problem in our setting because we are interested in how people's IAT changes when they are in a mixed room, and there is no particular reason for short-term extraneous factors to systematically differ across mixed and same race rooms.

Finally, Fiedler and Bluemke (2005) show that subjects who were instructed to fake an intended IAT outcome, using knowledge acquired about the test, were successful in doing so. At the time of our survey, IATs were not commonly used in South Africa and the freshmen to whom we administered the test did not report any familiarity with the test. Also, the IAT was the very first module in our survey, so students had not been primed in any way about the purpose of the research.

To conclude, the IAT has limitations, but most of these are not relevant in our setting, while the advantage of mitigating social desirability bias in self-reported attitudes is particularly important in the context of South Africa, where race issues are highly sensitive.

2.4 Attitudinal and behavioral measures

Through our survey questionnaire, we collected information on students' socioeconomic backgrounds, beliefs, friendships and attitudes towards other ethnic groups. More specifically, we elicited information about the following.

Friendship patterns. We asked: how often the respondent socialized with people of a different race in the past month; the self-reported preferred number of people of a different race in a hypothetical study group or a leisure group formed by 7 people; the share of (actual) best friends who are black or white; and the share of (actual) study-mates who are black or white.

Inter-racial attitudes. We asked about the frequency and comfort in discussing issues of race; agreement with abolishing affirmative action in university admission; feelings of self-consciousness in dancing with a person of another race and in dating a person of another race.

[19] Dasgupta and Greenwald (2011) show that exposure to pictures of admired black Americans or disliked white Americans reduced implicit bias against black individuals, as captured by the IAT, 24 hours after the exposure.

Prosocial behavior. We asked if the respondent was a member of any community service or volunteer organizations and how much money he/she gave to charities (excluding churches) in the last year. For the follow-up survey, we also collected an experimental measure of prosocial behavior running a prisoner's dilemma game. We employed the strategy method and revealed the racial identity of participants using photographs.[20] The payoffs were as follows: R50 each if both players chose 'Cooperate'; R40 each if both players chose 'Block'; and if one player chose 'Block' while the other chose 'Cooperate', the former earned R75 while the latter earned R25.

2.5 Academic Performance

To measure students' academic performance we rely on administrative data. First of all, we know the Admission Point Score that the student received based on his/her performance in standardized high school final exams. APS is the average of the best 6 grades in the last year of high school. Exams always include English plus five additional exams (excluding Life orientation) and are used to assess whether a prospective student meets the specific admission requirements of their chosen course of study. Exams are graded in numerical percentage form, with a maximum of 100 per exam (600 in total) corresponding to the following grades: 80-100% = A, 70-79% = B, 60-69% = C, 50-59% = D, 40-49% = E, 34-39% = F, 30-39% = FF, 20-29% = G, 0-19% = H. We rescale this variable dividing it by 1000 and denote it as '*UCT admission score*', using it as a proxy for students' ability at the beginning of their career at UCT.

We then have several measures of performance at the end of the first and second academic year, collected by the university's registry. These include the total number of exams passed and failed and their grade point average (GPA). In our analysis we employ the average GPA the student obtained, standardized to have mean 0 and standard deviation 1 in each year and we denote this variable as '*GPA*'. In addition to the GPA and the number of exams passed, we use a third indicator based on students' academic evaluation by the Faculty Examination Committees.[21] This indicator, denoted as '*Eligible to continue*', takes value one if the student is in good standing and eligible to continue studying in the next academic year, possibly subject to passing some makeup exams.

[20] Note that the photos used in the game do not include any of the subjects in our sample. Experimental instructions are described in Online Appendix B.

[21] The students' evaluation is conducted at the end of the academic year and takes into account the number of exams passed and their grades.

3 Descriptive statistics and randomization

Our working sample consists of those students successfully interviewed at baseline and follow-up with non-missing values for both IAT tests, which is 499 students. Of these, 157 were allocated to a roommate of a different race and 342 were sharing the room with a student of their own race. The racial composition of this sample is as follows: 332 respondents are black, 117 are white, 18 are coloured and 32 are Indian, Asian or other race. Notice that this composition mechanically generates differences in the probability of being in a mixed room for different races, with the more numerous group (black students) having lower probability of being in a mixed room.[22]

To benchmark our empirical shares against those generated by perfect random matching, we conducted some resampling-type tests. We took the shares of the various groups in each residence (restricted to first-year students and including freshmen who did not take our survey but were in the residence) and simulated random pairings based on 10,000 replications. We obtained an average probability of being in a mixed room equal to 0.52. In our sample the empirical probability is 0.31, significantly lower (p-value 0.001). This could raise concerns that the wardens deviated from the random allocation policy, which may invalidate our inference. To assess this point we bring a number of pieces of evidence showing that, despite the significant discrepancy, we believe the allocation can be considered as good as random.

First, Table 1 reports summary statistics at baseline for the main outcomes of interest and the controls for the full sample (columns 1-2), for students in mixed rooms (columns 3-4) and for those in non-mixed rooms (columns 5-6). The last two columns show the difference in means between non-mixed and mixed rooms and the associated p-value. In Panel A we include the full sample, while in Panels B and C we separately consider white and black students, respectively, since we will be running regressions on these subsamples.[23] Overall, none of the variables considered are statistically different between the two groups. Importantly, our two main outcomes of interest, the *Race IAT* and the *Academic IAT*, are well balanced between mixed and non-mixed rooms, both on aggregate and within racial groups. In the full sample, the means of the Race and Academic IAT scores are -0.19 and -0.21, respectively (recall that negative values of the IAT indicate prejudice against black relative to white individuals). When we split

[22] In our sample the fraction living with a roommate of a different race is .23 for black, .33 for white, .78 for Indian, Asian and other and 1 for coloured students, who all happen to be allocated to non-coloured roommates.

[23] Summary statistics at follow-up for the main variables are reported in Appendix Table A4.

by race, white students exhibit higher prejudice levels than black ones, though even the latter group is found to hold negative stereotypes against itself. Turning to other controls, it is noteworthy that the UCT admission score, a proxy for academic ability at baseline, is on average identical for students in mixed and non-mixed rooms in all three panels. Socioeconomic controls are also well balanced across treatment arms.

Second, as an additional check, in Table 2 we report the coefficients of a regression of the likelihood of being assigned to a mixed room on individual pre-treatment characteristics for the full sample and separately by race.[24] For our key outcome variables, the Race and Academic IAT scores, no evidence of sorting appears at baseline. This is also true for other controls when we look at white and black students separately. The only variable that is significant at the 5 percent level is the dummy for coloured respondents in the full sample (columns 1-2). The dummy for Indian/other is significant at the 10 percent level in column 1, not in column 2. To some extent, this pattern is to be expected: groups that have lower population shares in the dorms compared to black students (the omitted category) mechanically have a higher probability of getting a roommate of a different race.[25]

In panel B of Table 2, we look at the correlation between the probability of being in a mixed room and explicit inter-racial attitudes at baseline. Once again, the allocation to a mixed room is not significantly correlated with pre-treatment attitudes towards inter-group interaction.

In panel C of Table 2, we test for sorting by forming all $N * (N-1)$ possible dyads and estimating the probability that two individuals i and j are in the same room as a function of the distance in observables between i and j. In particular, we follow Fachamps and Gubert (2007) and Caprettini (2021) and estimate the undirectional dyadic regression:

$$SameRoom_{ij} = \alpha + \beta \left| X_i - X_j \right| + \gamma(X_i + X_j) + \varepsilon_{ij} \tag{1}$$

[24]Note that the number of observations in columns 3 and 5 (or 4 and 6) do not sum up to 499 because the full sample includes coloured students and other races. While we do not have a large enough sample to run separate estimations for these groups, when we pool them with black students the results are very similar to those reported in columns 5-6 (results available from the authors).

[25] Appendix Table A5 shows the results of a simulation exercise where we randomly assign roommates within dorms 10,000 times and compare the distribution of the simulated coefficients to the estimated coefficients of Table 2, Panel A. Specifically, the empirical p-values in columns 2 and 4 of Table A5 represent the fraction of simulations in which the simulated coefficient is smaller than the actual coefficient of Table 2. Based on this exercise, the only variable for which we detect significant deviation from random assignment is the dummy *coloured*.

where $SameRoom_{ij}$ takes value one if i and j are assigned to the same room, and X includes baseline values of the IAT scores, attitudes and socio-demographic characteristics. The coefficient of interest is β: a negative value of this coefficient would indicate sorting, i.e., that individuals who differ more in observables are less likely to be roommates.[26] Importantly, we show that differences in IAT scores do not predict being roommates, neither for the Race nor for the Academic IAT and there is no sorting on explicit attitudes either. We find some sorting on the UCT admission score (but not on attending private school) and on consumption (but not on wealth).

Finally, we perform two placebo tests. In Appendix Table A6, we run a regression using as dependent variables the lagged values of IAT scores and the *Index of Attitudes*, which are the outcomes we have for both baseline and follow-up (GPA and lab games are not available at baseline). We include as regressors our treatment variable and individual controls, and we show results with and without roommate's controls. The results show no evidence of sorting.

Overall, the results in tables 1 and 2 and the auxiliary tests performed increase our confidence that the roommate assignment mechanism can be considered as good as random for the purpose of our analysis.

4 Empirical strategy

To estimate the effects of exposure to a roommate of a different race on stereotypes, we estimate the following benchmark specification on the full sample and on the subsamples of black and white respondents:

$$Y_{ijkt} = \alpha Y_{ijk0} + \beta MixRoom_{ik0} + \boldsymbol{\gamma} Race_i + \lambda X_{ik0} + \mu X_{jk0} + \varphi RaceGrp_i * \delta_k + \varepsilon_{ijkt} \quad (2)$$

where Y_{ijkt} is the IAT score of student i paired with student j, in residence k, in the follow-up survey (time t) and Y_{ijk0} is the baseline (time 0) value of the same variable; $MixRoom$ is a dummy equal to 1 if at baseline the student was assigned a roommate of a different race; $Race_i$ is a vector of race dummies (White, Coloured, Indian or Asian or Other, with Black as omitted category); X_{ik0} is a set of individual controls measured at baseline that include: gender, UCT admission score, household wealth, the student's monthly consumption expenditure, and a dummy equal to one if the respondent is not

[26] The term $(X_i + X_j)$ is included to capture level effects and, as discussed in Fafchamps and Gubert (2007), its coefficient is not interpretable. For this reason, we do not display it in the table.

from South Africa.[27] The vector X_{jk0} is the same set of controls for the roommate; δ_k is a set of residence dummies, and ε_{ijkt} is the error term.[28] For the purpose of introducing race $\times$ residence fixed effects in the full sample specifications, we aggregate race groups into: Black, White, and a residual category – hence the variable $RaceGrp-$ due to the small number of observations in other groups.

Our coefficient of interest is β. A positive value of this coefficient indicates a reduction in prejudice against the black group (recall that negative values of IATs indicate negative stereotypes against black relative to white individuals, and positive values indicate stereotypes in favor of black ones).

For our other outcomes of interest, academic performance, attitudes and prosocial behavior, we estimate a specification analogous to (2) but without including the lagged dependent variable when not available (e.g., because academic grades or lab experiments are only measured at follow-up).[29] Also, in the specification where the dependent variable is academic performance, we add a set of dummies indicating the program in which the student is enrolled in. We estimate (2) using OLS with standard errors clustered at the room level. For those attitudinal variables that are categorical (and ordered), we employ an ordered logit model.

5 Results

5.1 Implicit Association Tests

Table 3 contains our first main result. It reports the estimated coefficients for equation (2) for the full sample and for white and black respondents separately. Standard errors in parentheses are clustered at the room level, and p-values in square brackets are adjusted for multiple hypothesis testing using Westfall and Young's (1993) free step-down resampling method for the family-wise error rate (FWER).[30]

[27] *Household wealth* is an index that measures per capita ownership of durable goods in the respondent's household. It is calculated by applying principal component analysis to the following categories of goods: computer, fridge, TV, landline and mobile phones, bicycles, motorbikes, bakkies, cars, electric stove, gas stove, kettle, and geyser. *Consumption* includes expenditure in the last month on lunches, dinners, food, alcohol, cigarettes, cell phone minutes and entertainment.

[28] To avoid restricting the sample due to missing values of some control variables, for control variables other than the baseline IAT score we replace missing values with the means and include in the regression dummies taking value one for those observations in which the missing value has been replaced.

[29] In the specifications where the dependent variable comes from the prisoner's dilemma game, we also include a dummy for whether the respondent knows the game player he/she has been matched with.

[30] We calculate the adjusted p-values with 10,000 replications. We first randomly assign individuals to a mixed or non-mixed room and then we assign them a roommate with the same residence and

The coefficient on *Mixed Room* captures the change in IAT for students who have been allocated a roommate of a different race, compared to students who have a roommate of their own race. In the full sample (column 1), results show that being exposed to a roommate of a different race has no significant effect on prejudice on average, as measured by the *Race IAT*. This is not surprising because, due to the way in which the IAT is defined (it captures relative prejudice towards one group versus the other), if each group became less prejudiced towards the other this would imply movements of the IAT in opposite directions. Indeed, when we split the sample between white and black respondents (columns 2 and 3), we find that exposure to a roommate of different race significantly reduces prejudice against black individuals for white respondents and has no significant effect for black ones. The magnitude of the effect for the white subsample (0.316) corresponds to a 0.63 standard deviation increase in the *Race IAT*, and indicates that exposure to a different race brings the average white respondent to the same level of prejudice as the average black respondent, which is quite noteworthy.[31] Although the psychology literature tends to consider the *Race IAT* as somewhat 'hard-wired', it has been shown to respond to stimuli that address inter-group bias.[32] In our setting, treatment consisted in daily, intense exposure to a member of the out-group for a full academic year, hence it is conceivable that it may have changed even deeply rooted stereotypes.

The coefficient on *Mixed Room* in the black subsample has the opposite sign (indicating a relative movement in favor of the other group) but is not significant. The lack of a significant effect on black respondents is not surprising in the context of psychological theories of intergroup contact. Pettigrew and Tropp's (2006) meta-analysis

gender that matches this characteristic (i.e., different race for a mixed room and same race for a non-mixed room). This allows us to keep the number of treated individuals constant in each iteration.

The outcome families differ from table to table, but within each table, a family consists of hypotheses with the same sub-sample. In Table 3 the p-values are adjusted for Race IAT and Academic IAT for the full, white and black subsamples separately. In Table 4 the p-values are adjusted for GPA, number of exams passed, and eligibility to continue. In Table 6 the family of outcomes includes the indexes of friendship, attitudinal measures and pro-social behavior. The results are similar if we use Anderson's (2008) False Discovery Rate (FDR) q-values (results available from the authors).

[31] To understand whether the reduction in white respondents' negative stereotypes is driven by being exposed to rommates with a relatively high academic quality, we tried interacting *Mixed Room* with the roommate's UCT admission score. The coefficient on the interaction term is insignificant (results available from the authors).

[32] For example, Rudman, Ashmore and Gary (2001) report evidence from two experiments showing that students enrolled in a prejudice and conflict seminar have lower implicit stereotypes after 14 weeks, compared to a control group. The seminar was taught by an African American professor, and involved students engaging in discussions on inter-group conflict.

found that the effect of interaction on the reduction of intergroup bias was significantly stronger for members of 'high-status' groups than for members of 'low-status' groups. While the literature reviewed by these authors mostly includes lab experiments and correlational studies, our findings confirm that this result holds in a field setting when we exploit exogenous variation in contact.

In columns 4-6, the dependent variable is stereotypes regarding academic performance, as measured by the *Academic IAT*. This variable captures the belief on the relative academic ability of black students versus white ones at the end of the first year, i.e., after the respondent has seen the performance of the roommate but also of other students at UCT. Results show that, on average, the simple exposure to members of another race does not change stereotypes on academic ability. In fact, we can reject that the effect of *Mixed Room* is the same on the *Race IAT* and on the *Academic IAT* using a Seemingly Unrelated Regression model (p-value 0.03).[33]

The results in Table 3 condition on roommate characteristics in order to isolate the effect of roommates' race. In Appendix Table A7, we replicate Table 3 without including roommate controls.[34] The estimates are very similar to those reported in Table 3, although the coefficient on *Mixed Room* in the white subsample drops by about 1/3 and is less precisely estimated. This suggests that part of the overall effect of exposure to a different group comes from the characteristics of that group.

In Appendix Table A8, we conduct a sensitivity analysis to show that our results are not driven by outliers, given the relatively small sample size. Once we exclude influential observations using the DFBETA method, the results become even stronger: the coefficient on *Mixed Room* for the sample of white students increases to 0.493 and becomes significant at the 1 percent level.[35]

Finally, in Appendix Table A9, we estimate the effect of roommate's race on IAT scores, separately by race of the roommate. We pool the three groups of coloured, Indian

[33] While we find effects on the 'taste-based' IAT (*Race IAT*) and not on ability stereotypes (*Academic IAT*), Beaman et al. (2009) find the opposite: exposure to female leaders reduced gender-occupation stereotypes, but left the association of gender with concepts of good and bad unaffected.

[34] When omitting roommate controls we estimate the overall effect of exposing someone to a person of a different race, including the differences in socioeconomic characteristics that come with it. This is a relevant policy benchmark for assessing the overall effect of mixing students from different backgrounds.

[35] The DFBETA method identifies observations that have a large influence on parameter estimates by re-estimating the model dropping one observation at a time and calculating a scaled measure of the change in the parameter of interest when the i-th observation is deleted. For parameter j, the formula is $DFBETA = \frac{b_j - b_{(i)j}}{s_{(i)}\sqrt{(\mathbf{X}'X)_{jj}}}$. We follow Belsley et al.'s (1980) recommendation and drop observations with $DFBETA > 2/\sqrt{N}$.

and Asian respondents due to the small sample size. We find that white respondents (col. 2), both rooming with a black and rooming with a coloured/Indian/Asian, reduce implicit bias, i.e., increase the *Race IAT*. The effect is more precisely estimated for black roommates, but we cannot reject that the two coefficients are the same (p-value 0.389). black respondents (col. 3), having non-black roommates, reduce the *Race IAT*, but not in a significant way. Again, we cannot reject the coefficients for white and coloured/Indian/Asian roommate being the same. For the *Academic IAT*, no significant effect is detected, as in Table 3.

5.2 Academic Performance

A central motivation underlying the policies of many universities that apply random assignment of roommates is to mix students who have different academic achievement. It is therefore important to investigate the effect of having a roommate of a different race on a student's academic performance.

Appendix Figure A3 shows the distribution of the average (standardized) GPA for black and white students at the end of the first academic year.[36] White students have a higher academic performance compared to black ones, and the gap is substantial: the mean GPA for white students is 0.59 while that for black ones is -0.23 (see Appendix Table A4).

In Table 4 we test whether having a roommate of a different race significantly affects this gap, using three different outcomes for academic performance. In columns 1 to 3 the dependent variable is the student's GPA at the end of the first year. Regressors (not shown) include all the controls of the benchmark specification (see Table 3) for the respondent and for the roommate, plus academic program fixed effects.[37] On average, students in mixed rooms do not have a significantly higher GPA than students in same-race rooms (column 1), but this effect masks an important heterogeneity. Black students who share the room with someone of a different race improve their GPA by 0.257 (column 3), corresponding to .27 of a standard deviation of the variable. The coefficient is significant at the 5 percent level even after adjusting for multiple hypothesis testing. The magnitude of the implied effect amounts to closing about 1/3 of the gap in the average GPA between white and black students. For white students, being in a mixed room has no significant impact on GPA (column 2): the estimated coefficient is

[36] The pattern when we pool black, coloured and other race students in a single category is very similar to that of black students alone.

[37] Academic programs are a subdivision of college tracks within each faculty. Within each academic program, students in the first year typically take the same (compulsory) courses.

very close to zero (−0.028, with a standard error of 0.243). While we have relatively low statistical power for the white subsample, the size of the effect for this group is one order of magnitude smaller than that for black students.[38]

In the remaining specifications of Table 4, we consider two additional measures of academic performance. The dependent variable in columns 4-6 is the number of exams passed during the first year (ranging from 0 to 9). Being allocated to a mixed race room increases the number of exams passed in the first year by .45 on average. This effect is particularly strong among black students: an increase of 0.65 exams (corresponding to 0.23 std. dev.), significant at the 5 percent level. For white ones, we find a small and insignificant negative impact (−0.17, corresponding to 0.07 std. dev.). The improvement for black students is noteworthy, as the difference in the number of exams passed between white and black students in same race rooms is quite striking: 2 exams.

In columns 7-9, the outcome we consider is *Eligibility to continue,* an assessment made by the Faculty Examination Committee based on the number of exams passed and the grades of each student. This variable takes value 1 if the individual is declared eligible to continue studying, possibly subject to passing makeup exams, and 0 otherwise. Once again, students in mixed rooms are more likely to be eligible to continue and this effect is driven by black students. For this subsample the estimated effect of 15.2 percentage points implies an 18 percent increase over the mean in same race rooms, significant at the 1 percent level. In this case the effect on white students is actually positive (0.05), albeit insignificant.

In the last three columns of Table 4 we estimate the effect of having a roommate of a different race on a summary index of performance, computed as the first principal component of *GPA*, *Number of exams passed* and *Eligibility to continue*. Being in mixed room is strongly positively correlated with the index of performance for black students (a coefficient of 0.44, corresponding to 0.38 std. dev. for this sample), while the impact for white ones is virtually zero.

Estimates without controls for roommate characteristics and sensitivity analysis excluding influential observations are reported in Appendix Tables A10 and A11, respectively. The results are very similar to those in Table 4.

Appendix Table A12 reports analogous estimates for academic performance in the second year. Given that students are free to change room –and to leave the residence

[38] Given our sample size, we can rule out a deterioration of more than 0.58 standard deviations and an improvement of 0.52 standard deviations in the GPA of white students.

altogether– in their second year, it is interesting to test whether the effects of exposure during the first year are persistent. Table A12 shows that the effect on GPA fades away, while the effects on the number of exams passed and on eligibility to continue remain significant in the full and in the black sample, and of similar magnitude (the effect is actually bigger for the number of exams passed, amounting to *one* extra exam).[39] This is remarkable because the outcomes in Table A12 do not cumulate performance in years 1 and 2, but only refer to year 2, i.e.: GPA in second-year exams, number of exams passed in the second year, and eligibility to continue to the third year.

What explains the impact of inter-group contact on performance? The first, obvious channel may be 'standard' academic peer effects. Since UCT applies a certain degree of affirmative action in admissions, black students have on average lower UCT admission scores compared to other groups, reflecting the legacy of unequal apartheid schooling systems by race. This implies that black students who have a non-black roommate are on average matched with someone who has a stronger academic background. Thus, one may conjecture that it is not exposure to a member of a different race, but exposure to a higher-achieving peer that generates the performance gains for black students.

This explanation is not supported by the data. First of all, the benchmark regressions in Table 4 directly control for roommate's ability (proxied by UCT admission score) and this variable is never significant.[40] Appendix Table A13 further corroborates this evidence. In columns 1 and 2, we control for whether a black student and their roommate are in the same faculty (30 percent of the students in our sample are). For freshmen in the same faculty it should be easier to study together or help each other with coursework. In fact, being in the same faculty significantly predicts that two roommates study together (column 6). However, in columns 1-2, the interaction of Mixed Room with an indicator for same faculty has a positive, but insignificant, coefficient. Similarly, having at least one course in common with a roommate does not have a significant effect on performance for black students in mixed rooms (the coefficient on the

[39] The lack of persistence of the gains in GPA may suggest that, in their second year, students who had been exposed to a roommate of a different race focus on staying on track more than on improving their grades in every exam.

[40] Although it may seem surprising, the result that roommate's baseline ability does not predict performance is not uncommon in the peer effects literature. For example, Sacerdote (2001) shows that freshman and senior year GPA are significantly predicted by one's own high school academic score, but do not respond to the roommate's score. A similar result is reported by Angrist (2014), who finds that roommate's SAT reasoning score is unrelated to a student's own GPA. Finally, Zarate (2021) shows that being assigned to a neighbour with high admission score in Peruvian boarding schools does not improve academic performance on average. For low-achieving girls the effect is actually negative, which the author attributes to a decrease in self-confidence.

interaction term in columns 3-4 is small and insignificant).[41] Finally, students in mixed rooms are neither less nor more likely to study with their roommate than students in same race ones (columns 5-6). These findings suggest that 'standard' academic peer effects stemming from exposure to a higher-achieving peer or from studying together do not account for the higher academic performance of black students in mixed rooms.

To gain further insights on the factors that correlate with gains in academic performance, we use the machine learning (ML) approach proposed by Chernozhukov et al. (2020), which provides an agnostic procedure to detect heterogeneous treatment effects ex post. This can give us a sense of the characteristics of the individuals who benefit the most from being in a mixed room. The details of our procedure are described in Online Appendix C. We use the ML proxy predictor of conditional average treatment effects (CATEs) for academic outcomes –for compactness we focus on the *Index of Academic Performance* and on *GPA*. We have mixed evidence on heterogeneous effects, with non-significant estimates for the Best Linear Predictor of CATE (see Appendix Table C1) but highly significant differences in the Classification Analysis (CLAN). This is consistent with an underpowered sample for the detection of slightly differential effects, but with evidence of heterogeneity based on the CLAN differences.

Table 5 reports the CLAN estimates for the main covariates of interest, i.e., the average characteristics of the 20% least affected (columns 1, 3) and 20% most affected (columns 2, 4) students on the basis of the ML proxy predictor. The analysis confirms that white students are over-represented in the least affected group, and are basically absent from the most affected one. We also find significant differences on a number of respondent characteristics, indicating that the individuals who benefit the most from being in a mixed room have lower UCT entry score, are less likely to come from a private school and are poorer both in terms of consumption and wealth. Interestingly, they do not have higher ability roommates –if anything, the opposite.[42] Importantly, the students who benefit the most from inter-group contact have roommates with significantly less negative stereotypes towards black individuals at baseline, both in terms of *Race IAT* and of *Academic IAT*. Although we cannot give a causal interpretation

[41] For each student in our dataset, we have administrative data on the list of courses they took in the first year, hence we can observe whether an individual and his/her roommate took the same course. Note that during freshman year the list of courses is essentially predetermined by the requirements of each program, so concerns of endogeneity in course choice should not affect our estimates.

[42] The fact that the roommate's UCT is lower in the most affected group may relate to improvements in students' self-image and a reduction in stereotype threat (e.g., Steele and Aronson, 1995). For a meta-analysis that is critical of the stereotype threat literature, see Flore and Wicherts (2015).

to this result, it is consistent with the idea that reducing racial stereotypes may bring significant benefits not only in terms of social cohesion, but also of academic outcomes. To the best of our knowledge, our analysis is the first to jointly document the effect of inter-group contact on prejudice reduction and academic performance, which seems very relevant for the debate on racial integration in universities.[43]

To further explore the mechanisms of interaction, in the next section we examine the effects of roommate race on a variety of attitudinal and behavioral measures.

5.3 Attitudinal and behavioral measures

Our final set of results test for the presence of behavioral changes induced by inter-group interaction that could be consistent with the stereotype reduction and academic gains documented above. We start from a set of 15 outcomes that include behavioral and attitudinal measures that we collected through our survey and through lab experiments. To avoid over-rejecting the null hypothesis due to multiple inference (Anderson, 2008), we first group the 15 outcomes into three main indices that aggregate information across related variables: we label them 'friendship', 'attitudes' and 'prosocial behavior'. Since the original outcomes are measured at different levels (some are binary, other ordinal, other continuous), we use as a summary index the first principal component computed using the polychoric correlation matrix of the outcomes in each group, which is theoretically invariant to changes in the number of levels. We also construct a global index of 'social behavior' that extracts the first principal component from all 15 outcomes at once. Since even with the summary indices we are still testing multiple hypotheses, we also report FWER-adjusted p-values in square brackets, applying Westfall and Young's (1993) resampling method with $10,000$ iterations. Table 6 reports these estimates; results for the disaggregated individual outcomes are reported in Appendix Tables A14, A15 and A16.[44]

Columns 1-3 of Table 6 report the effect of having a roommate of a different race on inter-racial friendships. The index of friendships includes the following variables: (i) the number of times the respondent socialized with someone of a different race in the

[43] Related but in different contexts, Carlana (2019) finds that female students perform less well in math when assigned to teachers with stronger gender stereotypes, and Glover et al. (2017) find that workers exert less effort when assigned to more prejudiced managers.

[44] The smaller sample size in Table 6 is due to some missing values in the underlying variables that enter the indexes. As robustness check, in Appendix Table A17 we replicate Table 6 replacing the missing values of each constituent variable with its mean, and adding among the controls corresponding dummies for missing values. This way the full sample is used in Table A17. We find very similar results both in terms of magnitude and statistical significance.

last month (excluding the roommate); (ii) the last time the respondent socialized with people of a different race;[45] (iii) the fraction of actual friends and study mates of a different race (excluding the roommate);[46] (iv) the number of desired friends of a different race in a hypothetical leisure group and in a hypothetical study group (excluding the roommate).[47] Column 1 shows that exposure to a roommate from a different racial group induces a sizeable increase in inter-racial friendships in the full sample, and this effect is significant at the 5 percent level after adjusting for multiple hypothesis testing. When we disaggregate by race of the respondent, the effect appears to be driven by white students, in line with our previous results for the *Race IAT*. Note that these effects are not mechanical given that we exclude the roommate from the friends or study mates. When we look at the individual variables composing the index of friendship (Appendix Table A14), we find that having a roommate of a different race generally has a positive impact on all components, for both white and black respondents, though the effect is typically less precisely estimated for the latter. These findings suggest that contact with a roommate of a different race leads to changes in the pattern of social interactions, which likely affects a range of individual outcomes, including stereotypes and academic performance.

In columns 4 to 6, we measure the effect of the treatment on an index of explicit attitudes. This index includes the following five variables: (i) the frequency with which the respondent talked about discrimination and racial bias with friends in the last month;[48] (ii) an indicator for whether the respondent feels comfortable talking to people about race and discrimination; (iii) a dummy taking value one if the respondent disagrees that affirmative action in university admissions should be abolished; (iv) an indicator for whether the individual disagrees with the statement: "I would probably feel a little self-conscious dancing with a person of another race in a public place"; and (v) an indicator for whether the individual disagrees with the statement "I would probably feel a little self-conscious having a girlfriend or boyfriend of another race". The coefficient on *Mixed Room* is positive and significant in the full sample (column 4), indicating that on average exposure to a roommate of a different race improved inter-racial attitudes. When we disaggregate by race of the respondent, again we find that the effect is driven

[45] Possible responses were: yesterday, last week, last month, last year, never.

[46] Friends were defined as "those you can turn to for help if needed" and we asked respondents to list the first name, gender, age and race of up to five friends.

[47] We asked our respondents how many people of different race they would want in a group of 7 people (including themselves), and they could choose a number between 0 and 6.

[48] Possible answers were: never, rarely, sometimes, most of the time, always.

by white respondents (column 5), consistent with the reduction in negative stereotypes by this group that emerged from our IAT results.

Columns 7 to 9 examine impacts on prosocial behavior, measured through an index that aggregates behavioral and experimental measures, in particular: (i) membership in community service or volunteer organizations; (ii) the amount of money given to a charity in the past year; and two experimental measures elicited through a prisoner's dilemma game: (iii) an indicator for whether the respondent chose to cooperate, and (iv) an indicator for whether the respondent believed that the partner would cooperate.[49] Our estimates suggest an improvement in prosocial behavior for all groups, but the effect is significant at the 10 percent level only for white students.

Finally, in the last three columns of Table 6, we aggregate all individual outcomes into a 'global' index of social behavior. We find positive and significant effects in the full and in the white sample, and a smaller (positive) insignificant effect in the black sample. The pattern for explicit attitudes and behaviors thus mirrors our earlier findings on implicit attitudes, where the strongest effects of inter-racial contact on stereotypes were found among white respondents.

As a last test of behavior change, we adopt a revealed preference approach and study residential choices for the students in our sample after the first year. If the students have opted to remain in the residences, through the administrative data we know in which residence and room they are and, if it is a double room, we know the identity of their roommate. If the students have left the university residence system, we do not have any information on their location.

Panel A of Table 7 shows that having been assigned a roommate of a different race in the first year is uncorrelated with the choice of staying in a university residence during the second year. Panel B shows insignificant effects also for the probability of being in a residence *and* in a mixed room in year 2. Finally, panel C shows that mixed room status in year 1 does not predict the probability of being in a residence and with the same roommate in year 2. These results are encouraging because, if inter-racial contact had generated significant 'unhappiness', we would expect students in mixed rooms to be more likely to opt out of the residence system altogether –or at least less likely to be in a mixed room or with the same roommate– in the second year, while we find that this is not the case.

[49] In our sample, 57 percent of the subjects chose to cooperate and 61 percent stated that they believed the partner would cooperate.

6 Conclusions

Social diversity does not always translate into equal opportunities for different groups in society: concerns exist about the possibility that today's schools, workplaces and social networks display significant segregation along income or racial lines. Integration policies have been proposed as a way of improving equality of opportunities but also of reducing inter-group prejudice and conflict. In this paper, we take advantage of a policy designed to randomly allocate roommates in some residences at University of Cape Town to investigate the effects of exposure to individuals of a different race on implicit and explicit attitudes, behavior and academic performance. To the best of our knowledge, ours is the first attempt to link implicit bias, as measured by the Implicit Association Test (IAT), with academic performance in the context of a real world policy setting.

Our results point to a number of positive effects from inter-racial contact generated through this policy. First, we find that living with a roommate of a different race during the first year reduces white students' negative stereotypes against black individuals, as measured by the IAT. This effect is quite remarkable because a number of transformation initiatives have happened in post-apartheid South Africa that have made inroads in reducing the salience of race. Yet, the interaction generated by the roommate allocation policy is able to (further) reduce prejudice. We also find significant positive effects on explicit attitudes towards the other race and on inter-racial friendships, again most pronounced for white respondents.

Finally, we show positive effects of inter-group contact on the academic performance of the negatively stereotyped group: black students in mixed rooms significantly improved their GPA, passed more exams and were more likely to be eligible to continue university. The latter two effects (but not the first) persisted into the second year of university, when students may have changed residential setting. Our heterogeneity analysis suggests that the individuals who benefited the most from inter-group contact in terms of academic performance tended to be poorer and to have less-prejudiced roommates. These results are potentially important for the literature on academic peer effects, which has focused on the potential benefits of integration in terms of exposure to a different set of skills. Our findings point to the importance of assessing the impact of integration policies on both attitudinal measures (e.g., stereotype reduction) and performance, given that the two may positively reinforce each other.

It is worth discussing to what extent our findings generalize beyond the context we

study. Our experiment takes place in South Africa, a place that is certainly at the higher end of the distribution if one considers its history of inter-racial conflict and the significance of racial stereotypes. If anything, we think that this should have reduced the ex ante probability of success of a policy like the one we study. First, many would assume that such stereotypes were so deeply ingrained that it would have been difficult to affect them; second, because a large number of 'transformation' initiatives have been put in place in South Africa since the end of apartheid, the marginal effect of the roommate allocation policy we study may have been assumed to be small. The fact that in-depth interaction with the other group can reduce negative stereotypes even in historically charged contexts is a very encouraging message in this respect.

A second issue to consider in assessing external validity is that our sample is made of university students who are certainly not representative of the entire South African population. On the one hand, these students are likely to be the political and economic elites of tomorrow's South African society, so that a change in their attitudes and behavior can bring a certain significance for the country as a whole. On the other hand, if we consider how this may have affected our estimates, the direction of a potential bias is not obvious. If these students are more 'malleable' than the general population, then our estimates would be an upper bound. But if they were already more open-minded and less reliant on preconceptions when they entered university, then there would have been a lower margin for the intervention to have an effect and the bias could be in the opposite direction.

Overall, we believe the specificity of the South African context should not be seen as a limitation of our results but, if anything, a strong proof of concept. More work to investigate the generalizability of our findings to different social and economic environments will help improve our ability to design integration policies in our increasingly diverse societies.

References

[1] Akerlof, G. and R. Kranton (2000), "Economics and Identity," *Quarterly Journal of Economics*, 115, 715–53.

[2] Alesina, A., M. Carlana, E. La Ferrara and P. Pinotti (2018), "Revealing Stereotypes: Evidence from Immigrants in Schools," NBER Working Paper 25333.

[3] Alesina, A., E. La Ferrara (2005), "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 43(3), 762-800.

[4] Allport, G.W. (1954), *The Nature of Prejudice*, Cambridge, MA: Perseus Books.

[5] Anderson, M.L. (2008), "Multiple Inference and Gender Differences in the Effects of Early Interventions: A re-evaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103 (484), 1481-95.

[6] Angrist (2014), "The Perils of Peer Effects" *Labour Economics*, 30, 98-108.

[7] Avitzour, E., A. Choen, D. Joel and V. Lavy (2020), "On the Origins of Gender-Biased Behavior: The Role of Explicit and Implicit Stereotypes," NBER Working Paper 27818.

[8] Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R. M., Harwood, J., Rubin, M., & Sibley, C. G. (2012), "The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice," *Personality and Social Psychology Bulletin*, 38, 1629-1643.

[9] Barnhardt, S. (2009), "Near and Dear? Evaluating the Impact of Neighbor Religion on Inter-Religious Attitudes: Experimental Evidence from India," mimeo, IFMR.

[10] Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande and P. Topalova (2009), "Powerful Women: Does Exposure Reduce Bias?," *Quarterly Journal of Economics*, 124(4), 1497-1540.

[11] Berge, L.I., K. Bjorvatn, S. Galle, E. Miguel, D. Posner, B. Tungodden and K. Zhang (2015), "How Strong Are Ethnic Preferences?," *NBER Working Paper* 21715.

[12] Bertrand, M., D. Chugh and S. Mullainathan (2005), "Implicit Discrimination," *American Economic Review*, 95(2), 94-98.

[13] Blanton, H., Jaccard J., Klick J., Mellers B., Mitchell G., Tetlock P. E. (2009),"Strong Claims and Weak Evidence: Reassessing the Predictive Validity of the IAT," *Journal of Applied Psychology,* 94, 567.

[14] Boisjoly, J., G. J. Duncan, M. Kremer, D. Levy, and J. Eccles (2006), "Empathy or Antipathy? The Impact of Diversity," *American Economic Review*, 96(5), 1890–1905.

[15] Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016), "Stereotypes," *Quarterly Journal of Economics*, 131(4), 1753-94.

[16] Caprettini, B. (2021), "Team Visibility and City Travel: Evidence from the UEFA Champions League Random Draw", *Journal of Sports Economics*, 22(1), 85-114.

[17] Carlana, M. (2019), "Implicit Stereotypes: Evidence from Teachers' Gender Bias," *The Quarterly Journal of Economics,* 134(3), 1163–1224.

[18] Carrell, S., B. Sacerdote and J. West (2013), "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 81(3), 855-882.

[19] Carrell, S., M. Hoekstra and J. West (2019), "The Impact of College Diversity on Behavior Toward Minorities," *American Economic Journal: Economic Policy,* 11(4), 159-82.

[20] Chernozhukov, Demirer M., Duflo E., Fernández-Val I. (2020), "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India," NBER Working Paper 24678.

[21] Dasgupta and Greenwald (2001) "On the Malleability of Automatic Attitudes: Combating Automatic Prejudice With Images of Admired and Disliked Individuals," *Journal of Personality and Social Psychology,* 81(5), 800-814.

[22] Duflo, E., P. Dupas and M. Kremer (2011), "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 101(5), 1739-74.

[23] Fafchamps, M. and F. Gubert (2007), "The Formation of Risk-Sharing Networks," *Journal of Development Economics*, 83(2), 326-350.

[24] Flore, P. C. and J. M. Wicherts (2015), "Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis," *Journal of School Psychology*, 53(1), 25-44.

[25] Fryer, R. and G. Loury (2013), "Valuing Diversity," *Journal of Political Economy,* 121(4), 747-774.

[26] Fiedler and Bluemke (2005), "Faking the IAT: Aided and Unaided Response Control on the Implicit Association Tests," *Basic and Applied Social Psychology*, 27(4), 307-316.

[27] Garlick, R. (2018), "Academic Peer Effects with Different Group Assignment Rules: Residential Tracking versus Random Assignment," *American Economic Journal: Applied Economics*, 10(3), 345-369.

[28] Glover D., A. Pallais and W. Pariente (2017), "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocy Stores," *Quarterly Journal of Economics*, 132(3), 1219-60.

[29] Green A. R., Carney D. R., Pallin D. J., Ngo L. H., Raymond K. L., Iezzoni L. I. and M. R. Banaji, (2007), "Implicit Bias among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients," *Journal of General Internal Medicine,* 22, 1231–38.

[30] Greenwald, A. G., & Banaji, M. R. (1995), "Implicit social cognition: Attitudes, self-esteem, and stereotypes," *Psychological Review*, 102, 4-27.

[31] Greenwald, A. G., & Banaji, M. R. & Nosek (2015), "Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects," *Journal of Personality and Social Psychology*, 108(4), 553–561.

[32] Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). "Measuring individual differences in implicit cognition: The Implicit Association Test," *Journal of Personality and Social Psychology*, 74, 1464-1480.

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

[34] Hanushek, Kain and Rivkin (2009), "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement," *Journal of Labor Economics*, 27(3), 349-83.

[35] Hasan, S., Bagde, S. (2013), "The mechanics of social capital and academic performance in an Indian college," *American Sociological Review,* 78(6), 1009-1032.

[36] Hoff, K., Pandey P. (2006), "Discrimination, Social Identity and Durable Inequalities," *American Economic Review*, 96(2): 206-211..

[37] Hjort, J. (2014), "Ethnic divisions and production in firms," *Quarterly Journal of Economics*, 129(4), 1899-1946.

[38] Kling, J.R., J.B. Liebman and L.F. Katz (2007), "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75(1), 83-119.

[39] Kurdi B., A. E. Seitchik, J. R. Axt, T. J. Carroll, A. Karapetyan, N. Kaushik, D. Tomezsko, A. G. Greenwald, M. R. Banajii (2018), "Relationship Between the Implicit Association Test and Intergroup Behavior: A Meta-Analysis," *American Psychology*, DOI: 10.31234/osf.io/582gh.

[40] Lowe M. (2021) "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration," *American Economic Review*, 111(6), 1807-44.

[41] Lowes, S., N. Nunn, J. A. Robinson and J. Weigel (2015), "Understanding Ethnic Identity in Africa: Evidence from the Implicit Association Test (IAT)," *American Economic Review Papers and Proceedings,* 105(5), 340-45.

[42] Lyle, D. (2009), "The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point," *American Economic Journal: Applied Economics*, 1(4), 69-84.

[43] McConnell and Leibold (2001), "Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes," *Journal of Experimental Social Psychology,* 37, 435–442.

[44] Mousa S. (2020), "Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq," *Science*, 369 (6505), 866-870.

[45] Oswald, F., Gregory M. L., Blanton H., Jaccard J., Tetlock P. E. (2013), "Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT Criterion Studies," *Journal of Personality and Social Psychology,* 105, 171.

[46] Paluck, E., S. Green and D. Green (2018), "The contact hypothesis re-evaluated," *Behavioural Public Policy*, 3(2), 1-30.

[47] Paolini, S., Harwood, J. and Rubin, M. (2010), "Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures", *Personality and Social Psychology Bulletin*, 36, 1723-1738.

[48] Pettigrew, T. and L. Tropp (2006), "A Meta-analytic Test of Intergroup Contact Theory," *Journal of Personality and Social Psychology*, 90(5), 751-783.

[49] Rao, G. (2019), "Familiarity Does Not Breed Contempt: Diversity, Discrimination and Generosity in Delhi Schools," *American Economic Review*, 109(3), 774-809.

[50] Rooth D. O. (2010), "Automatic associations and discrimination in hiring: Real world evidence," *Labour Economics,* 17 (3), 523-534.

[51] Reuben, E., Sapienza, P., & Zingales, L. (2014), "How stereotypes impair women's careers in science," *Proceedings of the National Academy of Sciences,* 111(12), 4403-4408.

[52] Rudman, L. A., Ashmore, R. D., and Gary, M. L., (2001). ""Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes," *Journal of Personality and Social Psychology*, 81(5), 856-868.

[53] Sacerdote, B. (2001): "Peer effects with random assignment: Results for Dartmouth roommates," *The Quarterly Journal of Economics*, 116(2), 681-704.

[54] Scacco, A. and S. Warren (2018), "Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria", *American Political Science Review*, 112(3), 654-677.

[55] Shayo, M. (2009), "A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution", *American Political Science Review*, 103(2), 147-174.

[56] Steele, C. and J. Aronson (1995), "Stereotype Threat and the Intellectual Test Performance of African Americans", *Journal of Personality and Social Psychology*, 69(5), 797-811.

[57] Stephan, W. G. and C. Stephan (1985), "Intergroup Anxiety", *Journal of Social Issues*, 41(3), 157-175.

[58] Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971), "Social categorization and intergroup behaviour," *European Journal of Social Psychology*, 1, 149-178.

[59] Uhlmann, E. L., Brescoll, V. L., & Paluck, E. L. (2006), "Are members of low status groups perceived as bad, or badly off? Egalitarian negative associations and automatic prejudice", *Journal of Experimental Social Psychology*, 42(4), 491-499.

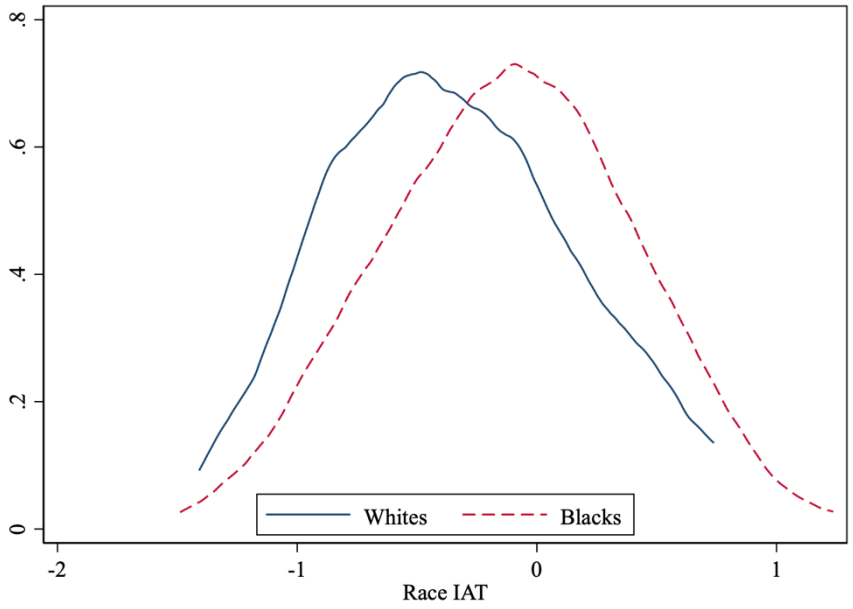
[60] Van Laar, C., S. Levin, S. Sinclair and J. Sidanius (2005), "The effect of university roommate contact on ethnic attitudes and behavior", *Journal of Experimental Social Psychology*, 41, 329–345.

[61] Westfall, P. and S. Young (1993), *Resampling-based Multiple Testing: Examples and Methods for P-value Adjustment*, John Wiley & Sons.

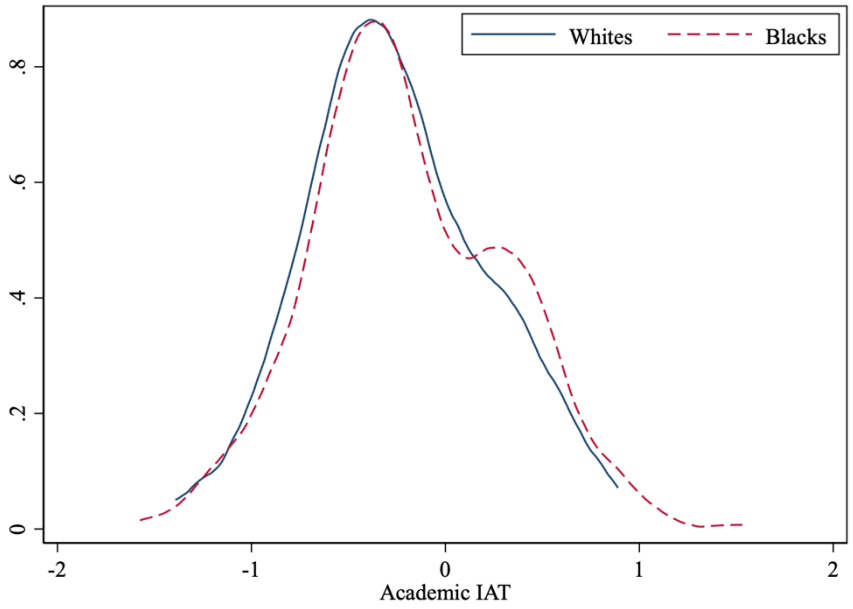
[62] Zárate, R. A. (2021), "Peer Effects on Social and Academic Skills," mimeo, University of Toronto.

Figures and Tables

Figure 1: Stereotypes as measured by IAT at baseline



Panel A: Race IAT



Panel B: Academic IAT

Note: Density of the Race and Academic IAT scores at baseline. Lower values of the score indicate more negative stereotypes towards blacks relative to whites.

Table 1: Descriptive statistics

	Full Sample		*Mixed Room*		*Non-Mixed Room*		*Mixed rooms - Non-mixed*	
	Mean	*Std. Dev.*	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*	*Beta*	*P-value*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Full Sample (N=499)								
Race IAT	-0.193	0.517	-0.252	0.496	-0.167	0.524	-0.057	0.308
Academic IAT	-0.213	0.494	-0.206	0.512	-0.216	0.486	0.048	0.374
UCT admission score	0.463	0.048	0.466	0.050	0.462	0.048	0.000	0.923
Wealth Index	0.024	2.122	0.045	1.890	0.014	2.223	-0.316	0.147
Consumption	0.926	0.847	1.021	0.929	0.882	0.804	0.083	0.343
Foreign	0.112	0.316	0.140	0.348	0.099	0.300	0.037	0.286
Private high school	0.601	0.490	0.618	0.487	0.594	0.492	-0.021	0.688
Panel B: Whites (N=117)								
Race IAT	-0.354	0.505	-0.318	0.465	-0.372	0.525	0.047	0.667
Academic IAT	-0.250	0.463	-0.229	0.488	-0.261	0.452	0.015	0.882
UCT admission score	0.487	0.039	0.487	0.042	0.487	0.038	0.003	0.725
Wealth Index	0.838	1.804	0.459	1.295	1.028	1.991	-0.613	0.103
Consumption	1.182	0.912	1.181	1.007	1.183	0.868	0.002	0.990
Foreign	0.068	0.253	0.103	0.307	0.051	0.222	0.067	0.210
Private high school	0.744	0.439	0.692	0.468	0.769	0.424	-0.079	0.390
Panel C: Blacks (N=332)								
Race IAT	-0.119	0.520	-0.174	0.538	-0.103	0.514	-0.081	0.239
Academic IAT	-0.188	0.498	-0.139	0.514	-0.203	0.494	0.062	0.341
UCT admission score	0.453	0.048	0.451	0.050	0.453	0.047	-0.001	0.833
Wealth Index	-0.380	2.051	-0.521	1.580	-0.339	2.170	-0.165	0.533
Consumption	0.809	0.800	0.894	0.901	0.784	0.769	0.130	0.204
Foreign	0.120	0.326	0.147	0.356	0.113	0.317	0.033	0.449
Private high school	0.536	0.499	0.533	0.502	0.537	0.500	0.001	0.983

Notes: The difference in col. (7) is the coefficient of the dummy *Mixed Room* in a regression that includes Race X Residence fixed effects in Panel A and Residence fixed effects in Panels B and C. *UCT admission score* is the sum of high school final grades, with weights depending on the specific department the student enrolls in; *Wealth index* measures per capita ownership of durable goods in the respondent's household and is constructed applying principal component analysis to the following categories of goods: computer, fridges, TV, landline and mobile phones, bicycles, motorbikes, bakkies, electricity, gas, kettles, geysers and cars; *Consumption* is the monthly consumption (in Rands) on lunches, dinners, food, alcohol, cigarettes, cell phone minutes, entertainment; *Foreign* is a dummy equal to one if the respondent is not from South Africa; *Private high school* is equal to one if the respondent was enrolled in a private high school before joining UCT.

Table 2: Probability of being in a mixed room at baseline

Sample:	*Full Sample*		*Whites*		*Blacks*	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Dependent variable = 1 if roommate of a different race at baseline						
Race IAT	-0.038	-0.021	0.075	-0.002	-0.068	-0.023
	(0.037)	(0.039)	(0.084)	(0.082)	(0.047)	(0.049)
Academic IAT	0.036	0.037	-0.019	-0.016	0.057	0.020
	(0.040)	(0.040)	(0.117)	(0.113)	(0.048)	(0.048)
White	-0.028	-0.071				
	(0.124)	(0.126)				
Coloured	0.627	0.647				
	0.234	0.256				
Indian/Other	0.432	0.459				
	0.239	0.262				
UCT admission score	-0.011	-0.013	0.494	0.858	-0.089	-0.106
	(0.446)	(0.421)	(1.214)	(1.254)	(0.520)	(0.494)
Foreign	0.092	0.090	0.199	0.118	0.083	0.058
	(0.078)	(0.073)	(0.198)	(0.196)	(0.090)	(0.084)
Private high school	-0.020	-0.018	-0.161	-0.141	0.001	0.003
	(0.042)	(0.041)	(0.112)	(0.120)	(0.050)	(0.049)
Wealth index	-0.016	-0.014	-0.031	-0.041	-0.013	-0.007
	(0.008)	(0.008)	(0.021)	(0.024)	(0.010)	(0.010)
Consumption	0.026	0.017	0.030	0.039	0.028	0.021
	(0.024)	(0.021)	(0.055)	(0.050)	(0.029)	(0.025)
R-squared	0.232	0.266	0.214	0.367	0.037	0.143
No. Obs.	499	499	117	117	332	332
Panel B: Dependent variable = 1 if roommate of a different race at baseline						
Index of Attitudinal measures	0.011	0.015	0.050	0.042	-0.004	-0.007
	(0.018)	(0.017)	(0.037)	(0.040)	(0.025)	(0.023)
R-squared	0.238	0.268	0.241	0.390	0.031	0.137
No. Obs.	455	455	112	112	295	295
Roommate controls	--	X	--	X	--	X
Residence FE	--	--	X	X	X	X
Race X Residence FE	X	X	--	--	--	--
Panel C: Dependent Variable = 1 if individuals i and j are in the same room						
Difference in Race IAT		0.000905			(0.000744)	
Difference in Academic IAT		0.000576			(0.000747)	
Difference in Index of Attitudinal measures		0.000307			(0.000348)	
Difference in UCT admission score		-0.000950			(0.000367)	
Difference in Wealth index		-0.000030			(0.000168)	
Difference in Consumption		-0.001330			(0.000588)	
Difference in Foreign		-0.002940			(0.00181)	
Difference in Private high school		-0.000794			(0.000685)	
No. Obs.			59,522			

Notes: In Panels A-B we report OLS estimates for the probability of being in a mixed room. Each observation is a respondent. Standard errors in parentheses are clustered at the room level. In Panel C, we report dyadic regressions for the probability that any two individuals within a residence are in the same room. Each observation is a pair of respondents. Standard errors in parenthesis are corrected for dyadic correlation following Fafchamps and Gubert (2007) and Caprettini (2021). The dyadic regression also includes among the controls the sum of each variable across individuals *i* and *j* and Residence fixed effects. Variables are defined in the footnote of Table 1.

Table 3: Stereotypes and exposure to a roommate of different race

Dependent variable:	*Race IAT*			*Academic IAT*		
Sample:	*Full Sample*	*Whites*	*Blacks*	*Full Sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)
Mixed Room	-0.008	0.316	-0.094	-0.023	0.014	-0.009
	(0.060)	0.140	(0.069)	(0.047)	(0.107)	(0.059)
	[0.892]	[0.055]	[0.319]	[0.864]	[0.992]	[0.884]
Controls[(a)]	X	X	X	X	X	X
Roommate controls[(b)]	X	X	X	X	X	X
Mean of dep.var. in same race room	-0.178	-0.423	-0.097	-0.209	-0.293	-0.185
R-squared	0.140	0.217	0.097	0.106	0.266	0.087
No. Obs.	499	117	332	499	117	332

Notes: OLS estimates with standard errors in parenthesis clustered at the room level. P-values in square brackets are adjusted for multiple inference using the resampling method of Westfall and Young (1993) with 10.000 interactions. Higher values of the dependent variable (IAT) indicate less prejudice against blacks. All control variables are measured at baseline. In the full sample (cols. 1 and 4) we control for the race of the respondent (White, Coloured and Indian/Others) with Black as omitted category and Race X Residence fixed effects. In cols. 2, 3, 5, 6 we control for Residence fixed effects. (a) Controls include IAT at baseline, female dummy, UCT admission score, wealth index, consumption, foreign, private high school, as defined in the footnote of Table 1. (b) Controls for roommate include: UCT admission score, wealth index, consumption, foreign and private high school (female not included, as all rooms are single sex).

Table 4: Impact on academic performance

Dependent variable:	*GPA*			*Number of exams passed*			*Eligible to continue*			*Index of Performance*		
	Full sample	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mixed Room	0.147	-0.028	0.257	0.447	-0.168	0.645	0.105	0.050	0.152	0.289	0.010	0.443
	(0.102)	(0.243)	(0.125)	(0.204)	(0.523)	(0.245)	(0.031)	(0.066)	(0.040)	(0.113)	(0.259)	(0.141)
	[0.145]	[0.929]	[0.040]	[0.046]	[0.929]	[0.017]	[0.002]	[0.789]	[0.001]			
UCT admission score	8.237	10.812	5.505	12.750	10.553	11.441	0.840	-0.453	0.738	8.105	7.702	6.158
	(1.204)	(2.687)	(1.392)	(2.346)	(5.881)	(2.841)	(0.387)	(0.634)	(0.469)	(1.320)	(2.090)	(1.618)
Roommate's UCT admission score	-0.098	-0.683	0.029	-0.129	-1.544	0.145	0.045	0.035	0.030	-0.028	-0.640	0.078
	(0.334)	(0.746)	(0.512)	(0.474)	(1.131)	(0.641)	(0.078)	(0.171)	(0.107)	(0.342)	(0.675)	(0.520)
Controls	X	X	X	X	X	X	X	X	X	X	X	X
Roommate controls	X	X	X	X	X	X	X	X	X	X	X	X
Academic program FE	X	X	X	X	X	X	X	X	X	X	X	X
Mean of dep. var. in same race room	-0.044	0.663	-0.268	4.977	6.500	4.506	0.871	0.923	0.852	-0.042	0.709	-0.281
R-squared	0.424	0.576	0.386	0.709	0.727	0.715	0.325	0.436	0.400	0.447	0.426	0.447
No. Obs.	499	117	332	499	117	332	498	117	332	498	117	332

Notes: OLS estimates with standard errors in parenthesis clustered at the room level. P-values in square brackets are adjusted for multiple inference using the resampling method of Westfall and Young (1993) with 10.000 interactions. The dependent variable in cols. 1-3 is the GPA (standardized over the full sample); in cols. 4-6 it is the number of exams passed during the first year; in cols. 7-9 it is a dummy for being in good standing and eligible to continue the following year; in cols. 10-12 it is an index constructed as the first principal component of the previous three variables. Controls and roommate controls are measured at baseline and described in footnote of Table 3. In the full sample we include the race of the respondent (White, Coloured and Indian/Other) with Black as the omitted category and Race X Residence fixed effects; in the white and black sub-samples we include Residence fixed effects. All regressions include Academic program fixed effects.

Table 5: CLAN of academic performance

Dep. var.:	*Index of Academic Performance*				*GPA*			
	20% least	*20% most*	*Difference*	*p-value*	*20% least*	*20% most*	*Difference*	*p-value*
	(1)	(2)	(2)-(1)		(3)	(4)	(4)-(3)	
Respondent's characteristics								
White	0.707	0.011	-0.707	0	0.761	0.011	-0.772	0
	(0.561, 0.83)	(-0.007, 0)	(-0.837, -0.557)		(0.636, 0.885)	(-0.007, 0)	(-0.889, -0.636)	
Female	0.815	0.554	-0.25	0.008	0.707	0.663	-0.043	0.665
	(0.688, 0.92)	(0.398, 0.689)	(-0.443, -0.062)		(0.561, 0.83)	(0.513, 0.791)	(-0.225, 0.152)	
UCT entry score	0.475	0.46	-0.015	0.09	0.468	0.48	0.014	0.059
	(0.463, 0.488)	(0.446, 0.473)	(-0.034, 0.002)		(0.453, 0.481)	(0.467, 0.493)	(-0.005, 0.031)	
Private high school	0.902	0.228	-0.674	0	0.728	0.565	-0.185	0.086
	(0.8, 0.982)	(0.097, 0.338)	(-0.823, -0.527)		(0.586, 0.849)	(0.42, 0.71)	(-0.364, 0.02)	
Wealth Index	0.696	-0.682	-1.371	0	0.313	0.392	0.055	0.65
	(0.173, 1.232)	(-1.224, -0.197)	(-2.129, -0.689)		(-0.108, 0.822)	(-0.38, 1.015)	(-0.745, 0.93)	
Consumption	1.353	0.711	-0.649	0	1.189	0.806	-0.375	0.031
	(1.033, 1.658)	(0.549, 0.875)	(-1.005, -0.282)		(0.904, 1.476)	(0.634, 0.971)	(-0.718, -0.033)	
Foreign	0.098	0.054	-0.054	0.402	0.054	0.163	0.109	0.08
	(0.005, 0.169)	(-0.007, 0.103)	(-0.158, 0.052)		(-0.016, 0.103)	(0.047, 0.257)	(-0.023, 0.234)	
Race IAT	-0.283	-0.243	0.018	0.846	-0.242	-0.242	0.015	0.87
	(-0.423, -0.122)	(-0.402, -0.091)	(-0.192, 0.229)		(-0.394, -0.091)	(-0.395, -0.088)	(-0.204, 0.221)	
Academic IAT	-0.266	-0.163	0.101	0.286	-0.27	-0.175	0.103	0.286
	(-0.399, -0.136)	(-0.311, -0.022)	(-0.098, 0.293)		(-0.408, -0.131)	(-0.316, -0.035)	(-0.092, 0.293)	
Roommate's characteristics								
UCT entry score	0.479	0.449	-0.031	0.001	0.476	0.459	-0.016	0.115
	(0.467, 0.492)	(0.435, 0.464)	(-0.049, -0.012)		(0.463, 0.488)	(0.444, 0.474)	(-0.036, 0.004)	
Private high school	0.709	0.608	-0.109	0.248	0.655	0.601	-0.043	0.665
	(0.588, 0.836)	(0.478, 0.743)	(-0.288, 0.076)		(0.524, 0.784)	(0.468, 0.736)	(-0.227, 0.147)	
Wealth Index	0.55	-0.095	-0.618	0.108	0.243	0.496	-0.011	0.925
	(0.015, 1.071)	(-0.67, 0.468)	(-1.362, 0.161)		(-0.231, 0.702)	(-0.196, 1.192)	(-0.806, 0.838)	
Consumption	1.136	0.909	-0.217	0.193	1.096	0.911	-0.176	0.268
	(0.869, 1.423)	(0.738, 1.077)	(-0.544, 0.108)		(0.829, 1.36)	(0.754, 1.06)	(-0.508, 0.133)	
Foreign	0.09	0.069	-0.022	0.728	0.088	0.08	-0.017	0.807
	(0.007, 0.172)	(-0.001, 0.14)	(-0.129, 0.097)		(0.005, 0.169)	(0.001, 0.145)	(-0.118, 0.104)	
Race IAT	-0.292	-0.055	0.252	0.009	-0.288	-0.058	0.232	0.01
	(-0.433, -0.165)	(-0.174, 0.07)	(0.054, 0.45)		(-0.427, -0.155)	(-0.196, 0.081)	(0.048, 0.424)	
Academic IAT	-0.325	-0.073	0.265	0.002	-0.246	-0.137	0.099	0.262
	(-0.437, -0.212)	(-0.2, 0.054)	(0.096, 0.439)		(-0.365, -0.124)	(-0.257, -0.013)	(-0.081, 0.273)	

Notes: Classification Analysis (CLAN) of Academic Performance Index (columns 1-2) and GPA (columns 3-4). The table reports the average characteristics of the 20% least affected (columns 1 and 3) and 20% most affected (columns 2 and 4) units defined in terms of the ML proxy predictor. Medians over 100 splits. 90% confidence intervals in parenthesis. P-values refer to the hypothesis that the difference is equal to zero.

Table 6: Impact on friendships, explicit attitudes and pro-social behaviour

Dependent variable:	*Index of friendship*			*Index of explicit attitudes*			*Index of pro-social behavior*			*Global Index of social behavior*		
Sample:	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mixed Room	0.340	0.477	0.254	0.318	0.671	0.072	0.169	0.438	0.229	0.439	0.760	0.196
	(0.128)	(0.187)	(0.170)	(0.126)	(0.261)	(0.166)	(0.138)	(0.250)	(0.165)	(0.150)	(0.294)	(0.212)
	[0.032]	[0.050]	[0.363]	[0.032]	[0.050]	[0.673]	[0.233]	[0.094]	[0.363]			
Controls	X	X	X	X	X	X	X	X	X	X	X	X
Roommate controls	X	X	X	X	X	X	X	X	X	X	X	X
Mean of dep.var. in same race room	-1.110	-1.053	-1.173	-0.810	-1.642	-0.562	-0.705	-0.873	-0.651	-1.457	-1.604	-1.465
R-squared	0.317	0.505	0.155	0.186	0.369	0.068	0.168	0.374	0.099	0.321	0.458	0.149
No. Obs.	411	94	275	453	106	299	388	94	253	315	79	203

Notes: OLS Estimates with standard errors in parenthesis clustered at the room level. P-values in square brackets are adjusted for multiple inference using the resampling method of Westfall and Young (1993) with 10.000 interactions. The indexes used as dependent variables are constructed using a polychoric principal component analysis. The *Index of friendship* (cols.1-3) includes the following variables: (i) no. of times respondent hung out with people of different race in the last month: =0 if never. =1 if once. =2 if 2-5 times. =3 if 5-10 times. =4 if more than 10 times; (ii) last time respondent hung out with people of different race: =0 if never. =1 if last year. =2 if last month. =3 if last week. =4 if yesterday; (iii) fraction of friends and study mates of a different race (excl. roommate); (iv) Preferred number of people of different race in leisure group and academic group. The *Index of explicit attitudes* (cols.4-6) includes: (i) "In the last month. how often did you talk with any friends of yours about topics of discrimination. prejudice and racial bias?": =0 if never. =1 if rarely. =2 if sometimes. =3 if most of the time. =4 if always; (ii) a dummy for whether respondent is comfortable talking about race; (iii) a dummy for whether respondent does not agree that affirmative action in University admission should be abolished; (iv) a dummy for whether respondent do not feel conscious dancing with a person of another race; (v) dummy for whether respondent do not feel conscious having a boyfriend/girlfriend of another race. The Index of *pro-social behavior* (cols.7-9) includes: (i) member of community service or volunteer organization; (ii) amount of money given to charity in the past year; (iii) dummy for whether respondent cooperated in the prisoner dilemma game; (iv) dummy for whether respondent believed partner would cooperate in prisoner dilemma. The *Global Index of social behavior* (cols.10-12) includes all the variables listed for the previous three indexes. In the full sample we control for the race of the respondent (White, Coloured and Indian/Other) with Black as the omitted category and Race X Residence fixed effects. In the white and black sub-samples we include Residence fixed effects. All regressions include individual controls and roommate controls as described in the footnote of Table 3.

Table 7: Residential choices at the end of the first year

Panel A: Dependent Variable = Still in Residence in year 2			
	(1)	(2)	(3)
Sample:	*Full sample*	*Whites*	*Blacks*
Mixed Room	0.073	0.014	0.061
	(0.048)	(0.110)	(0.055)
Controls	X	X	X
Roommate controls	X	X	X
Mean of dep.var. in same race room	0.763	0.744	0.786
R-squared	0.118	0.134	0.084
No. Obs.	499	117	332
Panel B: Dependent Variable = Still in Residence and in Mixed Room in year 2			
Mixed Room	0.039	0.084	0.009
	(0.029)	(0.058)	(0.028)
Controls	X	X	X
Roommate controls	X	X	X
Mean of dep.var. in same race room	0.020	0.013	0.023
R-squared	0.052	0.158	0.056
No. Obs.	499	117	332
Panel C: Dependent Variable = Still in Residence and Same Roommate in year 2			
Mixed Room	0.019	0.038	0.023
	(0.027)	(0.037)	(0.026)
Controls	X	X	X
Roommate controls	X	X	X
Mean of dep.var. in same race room	0.020	0.013	0.023
R-squared	0.052	0.158	0.056
No. Obs.	499	117	332

Notes: OLS estimates with standard errors in parenthesis clustered at the room level. All variables are measured at baseline. Controls and roommate controls are described in footnote of Table 3.

Online Appendix
Interaction, Stereotypes and Performance. Evidence from South Africa

Lucia Corno, Eliana La Ferrara, Justine Burns

March 2022

A. Implicit Association Tests procedure

We implemented two versions of Greenwald et al.'s (1998) Implicit Association Test (IAT). In the first version, that we denote as 'Race-IAT', we asked respondents to match words or pictures of black and white students with positive and negative attributes. In the second version, denoted as 'Academic IAT', we asked them to match grades or pictures of black and white students with percentiles of the grade distribution. The procedure for each IAT can be described as follows.

Race IAT

Each respondent was invited to complete seven blocks, following the scheme in Figure A1. Examples of the screenshots of the various tasks are displayed in Figure A2.

- Block 1: The respondent was asked to categorize stimuli into two categories, Black-South Africans and White-South Africans, which appeared in the top left-hand and top right-hand corner of the screen, respectively. Pictures of a Black or a White person appeared one at a time at the center of the screen and respondents were instructed to sort each picture into the appropriate category as fast as possible by pressing the left-hand or the right-hand key.

- Block 2: The respondent had to complete a similar sorting task with a positive/negative attribute. The word "Positive" and "Negative" appeared in the top left-hand and top

right-hand corners, respectively, and a series of pleasant or unpleasant words appeared in the middle of the screen (see Figure A1 for the list of words). The respondent had to sort each word as being either positive or negative by hitting the left or right key.

- Blocks 3-4: The respondent had to perform combined tasks that included both the categories and attributes from the first two tasks. The combination "Black South-African/Positive" appeared in the top left corner and "White South-African/Negative" in the top right. Respondents would then see a series of stimuli in the middle of the screen, consisting of either pictures or words and had to press the left or the right-hand key depending on which category the picture/word belonged to.

- Block 5: Same as Block 1, but the position of "Black South-African" and "White South-African" was inverted.

- Blocks 6-7: Repetition of tasks 3-4, but with opposite pairings, that is, "White South-African/Positive" in the top left corner and "Black South-African/Negative" in the top right one.

  A score is produced at the end of the procedure, following the improved algorithm of Greenwald et al. (2003). This score reflects the difference in the reaction times in blocks 3-4 and 6-7. If race is differently associated with the attributes proposed (positive/negative), then it is expected that the pairing that a respondent implicitly associates will take less time. The final score is captured by our variable 'Race IAT', which takes higher values the lower the negative stereotype versus blacks relative to whites.

Academic IAT

The test for the 'Academic IAT' had the same structure as the 'Race IAT', except that the associations of race were with percentiles of the grade distribution instead of positive/negative words. The list of tasks, categories and screenshots are displayed in Figures A1 and A2. Here a brief outline of the procedure:

- Block 1: same as in the Race IAT

- Block 2: "First class (>75%)" appeared in the top left corner and "Third class (<60%)" in the top right corner. In the grading system at University of Cape Town, a grade of 75-100% First Class corresponds to an A in the US, and a grade of 50-59% Third

class corresponds to a C. In the middle of the screen a series of percentages appeared, ranging from 50% to 99%, and had to be classified in one of the two corners.

- Blocks 3-4: The combination "Black South-African/First class (>75%)" appeared in the top left corner and "White South-African/Third class (<60%)" in the top right. Respondents would then see a series of stimuli in the middle of the screen, consisting of either pictures or percentage grades, and had to press the left or the right-hand key depending on which category the picture/grade belonged to.

- Block 5: Same as Block 1, but the position of "Black South-African" and "White South-African" was inverted.

- Blocks 6-7: Repetition of tasks 3-4, but with opposite pairings, that is, "White South-African/First class (> 75%)" in the top left corner and "Black South-African/Third class (<60%)" in the top right one.

A score is produced at the end of the procedure, in an analogous way to the Race IAT. The resulting variable, 'Academic IAT', takes higher values the lower the association between blacks and poor academic performance, relative to whites.

B. Prisoner dilemma experimental instructions

In order to examine the impact of racial identity on exchange and cooperation, a series of prisoner dilemma games in which the racial identity of participants is revealed using photographs, were conducted during the follow-up survey in September 2012 among all the students who participated in the baseline survey.

We implemented a standard prisoners dilemma task where two students are paired and randomly assigned to their position as player A and player B. Each player sees a photograph of their partner. In this task, each player must choose whether to Cooperate with or Block their partner. The final payment in this task depends both on the choice that player B makes, as well as the choice made by player A. If both players choose "Cooperate", both will earn R50 each. If both players choose "Block", both will earn R40 each. If one player chooses Block while the other chooses cooperate, then the Player who chooses "Block"will earn R75, and the Player who chooses "Cooperate" will earn R25.

Here the exact instructions:

"This is a new decision-making task. Please read the instructions carefully, as it may differ from any previous decision-tasks you have participated in. In this decision-making task, you

are Player B. You are paired with another person, Player A. You will find a photograph of Player A on the next page. Your position as Player B was randomly assigned. In this task, must choose whether to Co-operate with Player A, or to Block Player A. Player A faces the same decision, and must decide whether to Co-operate with you, or to Block you. Your final payment in this task will depend both on the choice that you make, as well as the choice made by Player A. The table below describes the possible outcomes.

Possible choices	**Player A payment**	**Player B payment**
Both cooperate	50	50
Both block	40	40
A chooses Block/ B chooses Cooperate	75	25
A chooses Cooperate/B chooses Block	25	75

- If both players choose co-operate, both will earn R50 each.
- If both players choose Block, both will earn R40 each.
- If one player chooses Block while the other chooses co-operate, then the Player who chooses "Block" will earn R75, and the Player who chooses "Co-operate" will earn R25.

Any decision you make will be private. There are no tricks in this task. Everything is exactly as it has been described to you."

C. Heterogeneity Analysis

To assess treatment heterogeneity, we follow the generic machine learning (ML) inference approach proposed by Chernozukhov et al (2020), which provides an agnostic procedure to detect heterogeneity ex post. The procedure circumvents the obstacles posed by sparsity requirements in methods such as Athey and Wager (2018) by focusing on the features of the conditional average treatment effects (CATEs) instead of CATEs themselves. Based on these estimates, we focus on two key features, the Best Linear Predictor (BLP) of the CATE and the Classification Analysis (CLAN).

The variables Z used in the heterogeneity analysis are the following: Black; White; Coloured; Indian/Other; Female; UCT admission score; Foreign; Private high school; Wealth index; Consumption; Population IAT; Academic IAT; Index of Attitudinal measures; Preferred number of people of different ethnicity in a leisure group; Preferred number of people of different ethnicity in a leisure group; Talks about race; Is comfortable talking about race; Roommate's UCT admission score; Roommate is foreign; Roommate went to private school; Roommate's wealth index; Roommate's consumption; Roommate's Population IAT; Roommate's Academic IAT; Roommate's Index of Attitudinal measures; Feels discriminated (0-10); Expectations about top students in the program; Relative status of one's family; Importance of promoting racial understanding; Importance of helping others in difficulty; Importance of eliminating discrimination; importance of participating in civil rights organizations; Has a part-time job; Member of a sports team; Member of a music group.

Following the steps of the algorithm described in Chernozukhov et al. (2020), we implement the analysis as follows.

- **Step 0:** As in the main example of Chernozukhov et al. (2020), we fix the number of splits S = 100 and the significance level $\alpha = 0.05$.

- **Step 1:** Fix the propensity scores $p(Z)$, which in our case are calculated directly from the RCT design. Specifically, the propensity scores are the shares of treated observations within each *Race* x *Residence* bin (e.g., the share of white students in mixed rooms among all white students in residence r will represent the propensity score of white students in that residence).

- **Step 2:** Fix the proportion of the splits at 50% (half of our splits form the auxiliary sample N and the remaining half forms the main sample M). Each split follows these steps:

– **A:** Each of the selected ML methods is tuned and trained separately (in our case, the possible ML methods are LASSO, random forest and support-vector machine) to estimate $B(Z)$ and $S(Z)$, proxy predictors of b_0 and s_0, which, given our outcome variable Y and the dummy D representing the random allocation to a mixed-race room, are given by:

$$Y = b_0(Z) + Ds_0(Z) + U, \quad \mathrm{E}[U \mid Z, D] = 0$$

$$b_0(Z) = \mathrm{E}[Y \mid D = 0, Z]$$

$$s_0(Z) = \mathrm{E}[Y \mid D = 1, Z] - \mathrm{E}[Y \mid D = 0, Z]$$

– **B:** Estimate the BLP parameters by weighted OLS in the main sample M:

$$Y_i = \widehat{\alpha}' X_{1i} + \widehat{\beta}_1 \left(D_i - p\left(Z_i\right)\right) + \widehat{\beta}_2 \left(D_i - p\left(Z_i\right)\right)\left(S_i - \mathbb{E}_{N,M} S_i\right) + \widehat{\epsilon}_i, \quad i \in M$$

such that:

$\mathbb{E}_{N,M}\left[w\left(Z_i\right)\widehat{\epsilon}_i X_i\right] = 0$ for $X_i = \left[X_{1i}', D_i - p\left(Z_i\right), \left(D_i - p\left(Z_i\right)\right)\left(S_i - \mathbb{E}_{N,M} S_i\right)\right]'$, where $w\left(Z_i\right) = \left\{p\left(Z_i\right)\left(1 - p\left(Z_i\right)\right)\right\}^{-1}$.

Specifically regarding X_i, besides $B(Z)$, we also add the controls used in our main specification, which include academic program fixed effects, *Residence* fixed effects, *White* x *Residence* and *Black* x *Residence* fixed effects.

– **C:** Estimate the GATES parameters by weighted OLS in the main sample M:

$$Y_i = \widehat{\alpha}' X_{1i} + \sum_{k=1}^{K} \widehat{\gamma}_k \cdot \left(D_i - p\left(Z_i\right)\right) \cdot 1\left(S_i \in I_k\right) + \widehat{\nu}_i, \quad i \in M$$

Where ℓ_k is the (k/K)-quantile of $\{S_i\}_{i \in M}$. Once again, X_i includes the same controls as the ones described in step B.

– **D:** Estimate the CLAN parameters in the main sample M:

$$\widehat{\delta}_1 = \mathbb{E}_{N,M}\left[g\left(Y_i, Z_i\right) \mid S_i \in I_1\right] \quad \text{and} \quad \widehat{\delta}_K = \mathbb{E}_{N,M}\left[g\left(Y_i, Z_i\right) \mid S_i \in I_K\right]$$

where $I_k = [\ell_{k-1}, \ell_k)$ and ℓ_k is the (k/K)-quantile of $\{S_i\}_{i \in M}$.

– **E:** Compute the two performance measures for the ML methods:

$$\widehat{\Lambda} = \left|\widehat{\beta}_2\right|^2 \widehat{\mathrm{Var}}(S(Z)) \quad \widehat{\bar{\Lambda}} = \frac{1}{K}\sum_{k=1}^{K} \widehat{\gamma}_k^2$$

- **Step 3:** Choose the best ML methods based on the medians of $\widehat{\Lambda}$ and $\widehat{\bar{\Lambda}}$.
- **Step 4:** Compute the estimates, $(1-\alpha)$-level conditional confidence intervals and conditional p-values for all the parameters of interest.
- **Step 5:** Compute the adjusted $(1-2\alpha)$-confidence intervals and adjusted p-values using variational estimation and inference methods (VEIN). These methods take into consideration the two different sources of sampling uncertainty, that is, (i) the estimation uncertainty regarding our estimated parameters, conditional on the data split; (ii) the uncertainty or 'variation' induced by the data splitting (our split into the auxiliary N and the main M sample).

The algorithm described above is processed for two different outcomes, the *Academic Performance Index* and the *GPA*, for which the best learner based on the presented performance measures is the LASSO. The BLP results are provided below. The estimate for the mean prediction $\hat{\beta}_1$ (corresponding to the Average Treatment Effect) seems to match the estimated coefficient of our main analysis, and is statistically significant at the 10 percent level for the *Academic Performance Index*. However, the evidence of heterogeneity is somewhat mixed. Despite presenting positive point estimates for the differential prediction $\hat{\beta}_2$, none of the coefficients is statistically significant. An analysis of the CLAN results shows some consistency, with significant differences in the characteristics of the most affected by the intervention, who seem to have a higher share of black students, lower UCT scores, lower likelihood of coming from private schools and roommates with less negative stereotypes (see Table 5 in the text).

Online Appendix Figures and Tables

Figure A1: Structure of IAT

Block	Left key	Right key
1	Black South African	White South African
2	Positive words	Negative words
3	Black South African Positive words	White South African Negative words
4	Black South African Positive words	White South African Negative words
5	White South African	Black South African
6	White South African Positive words	Black South African Negative words
7	White South African Positive words	Black South African Negative words

Note: table refers to the Race IAT. The Academic IAT has the same structure, but "Positive words" is replaced by "First class (>75%)" and "Negative words" is replaced by "Third class (<60%)".

Items	Stimuli
Black South African	Picture of Black South Africans
White South African	Picture of White South Africans
Positive words	Good, joy, love, peace, wonderful, pleasure, glorious, laughter, happy
Negative words	Bad, agony, terrible, horrible, nasty, evil, awful, failure, hurt
First class (>75%)	99%, 85%, 78%, 90%, 82%, 95%
Third class (<60%)	52%, 56%, 55%, 50%, 57%, 59%

Figure A2: IAT screens

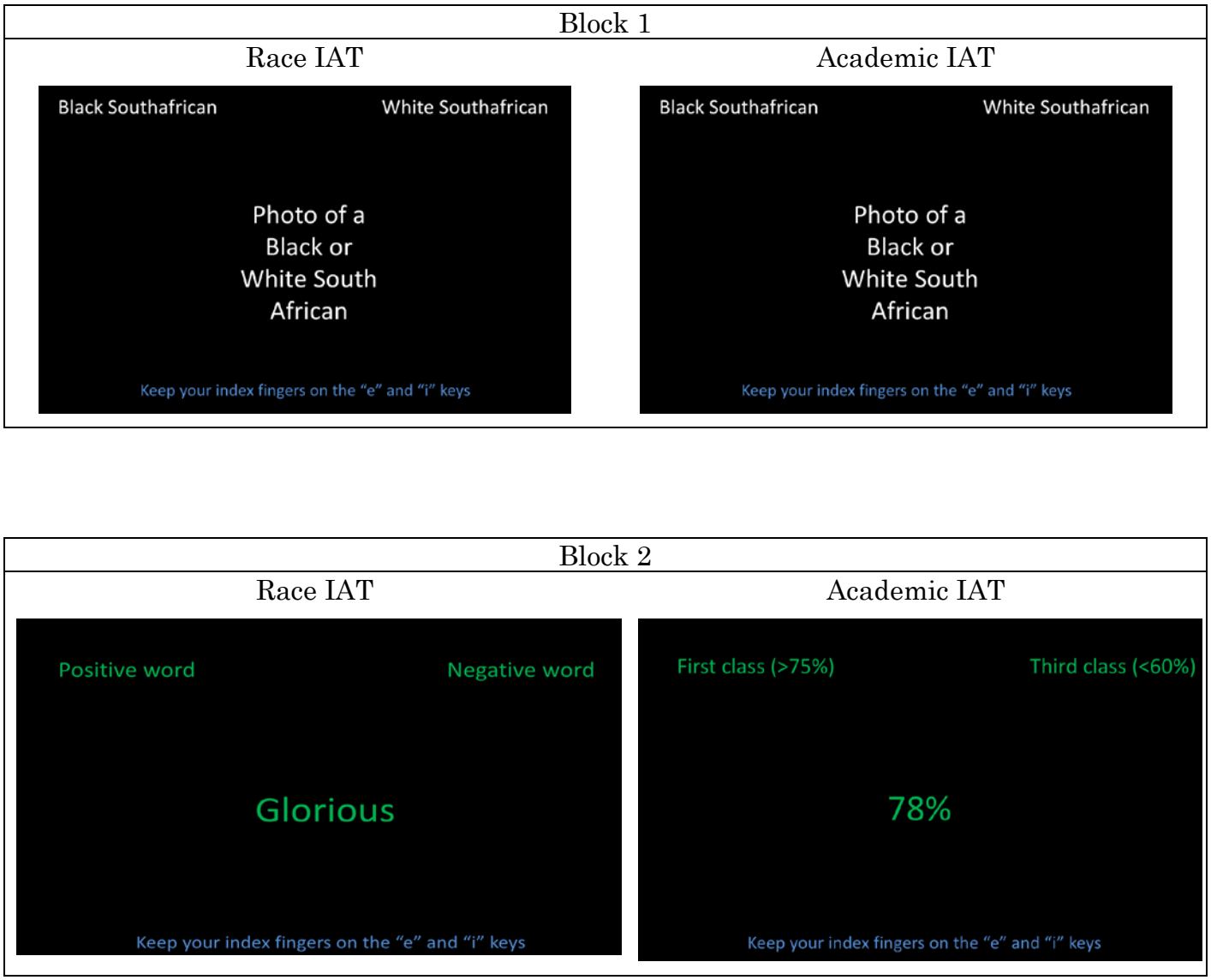
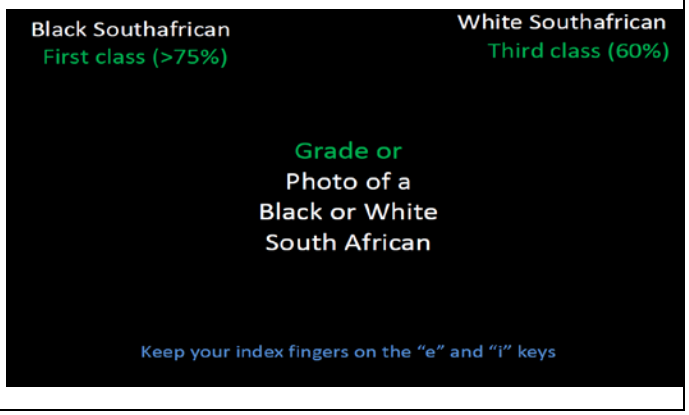


Figure A2 (cont'd)

Blocks 3 and 4

Race IAT

Academic IAT



Block 5

Race IAT

Academic IAT

Blocks 6 and 7

Race IAT

Academic IAT

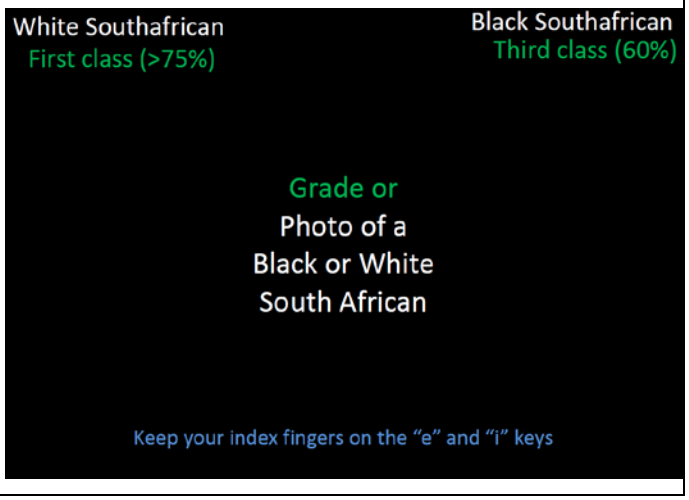


Figure A3: GPA distribution, by race

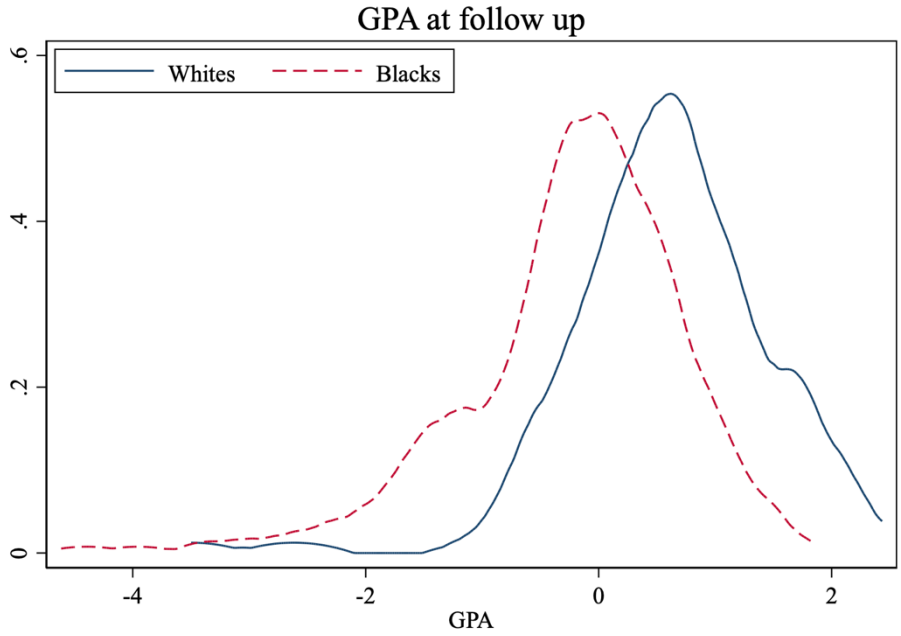


Table A1: Correlates of attrition

Dependent variable = 1 if respondent participated in follow-up survey							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mixed Room	-0.009			-0.007	0.004	-0.009	-0.011
	(0.043)			(0.044)	(0.044)	(0.043)	(0.043)
Race IAT		0.005		0.002		0.003	
		(0.031)		(0.036)		(0.036)	
Academic IAT			0.023		0.002		0.022
			(0.031)		(0.036)		(0.035)
Mixed Room*Race IAT				0.010			
				(0.066)			
Mixed Room*Academic IAT					0.066		
					(0.065)		
White*Race IAT						0.047	
						(0.076)	
Coloured*Race IAT						-0.201	
						(0.185)	
Indian/Other*Race IAT						-0.022	
						(0.209)	
White*Academic IAT							0.005
							(0.075)
Coloured*Academic IAT							0.104
							(0.156)
Indian/Other*Academic IAT							-0.090
							(0.108)
White	0.007	-0.001	0.002	0.008	0.029	0.024	0.016
	(0.340)	(0.337)	(0.336)	(0.338)	(0.335)	(0.331)	(0.340)
Coloured	-0.123	-0.128	-0.135	-0.119	-0.125	-0.171	-0.041
	(0.144)	(0.147)	(0.149)	(0.146)	(0.152)	(0.157)	(0.171)
Indian/Other	0.043	0.041	0.031	0.048	0.038	0.030	0.060
	(0.117)	(0.116)	(0.117)	(0.119)	(0.125)	(0.165)	(0.121)
Female	-0.138	-0.140	-0.139	-0.137	-0.137	-0.137	-0.136
	(0.054)	(0.054)	(0.053)	(0.055)	(0.055)	(0.055)	(0.055)
UCT admission score	0.332	0.333	0.351	0.334	0.348	0.291	0.335
	(0.378)	(0.379)	(0.381)	(0.380)	(0.382)	(0.381)	(0.383)
Foreign	0.091	0.090	0.090	0.090	0.095	0.095	0.094
	(0.057)	(0.056)	(0.057)	(0.057)	(0.057)	(0.057)	(0.057)
Private high school	-0.026	-0.026	-0.026	-0.027	-0.026	-0.028	-0.027
	(0.035)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)
Wealth index	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.004
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Consumption	-0.046	-0.046	-0.046	-0.046	-0.046	-0.046	-0.046
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Mean of dep.var.	0.804	0.804	0.804	0.804	0.804	0.804	0.804
R-squared	0.126	0.126	0.126	0.126	0.128	0.129	0.128
No. Obs.	621	621	621	621	621	621	621

Notes: OLS estimates with standard errors in parenthesis clustered at the room level. Controls variables are measured at baseline and include Residence X Race fixed effects. All the controls are measured at baseline.

Table A2: Comparison of our sample with other UCT students

	Our sample		*Students in our residences (single & double rooms) not in the sample*			*All first year students (residence & non residence) not in the sample*			*All students in all years (residence and non residence) not in the sample*		
	N	*Mean* (1)	*N*	*Mean* (2)	*P-value* (2)-(1)=0	*N*	*Mean* (3)	*P-value* (3)-(1)=0	*N*	*Mean* (4)	*P-value* (4)-(1)=0
Panel A: Whites											
Female	117	.675	444	.707	.502	844	.495	.000	1842	.518	.001
UCT admission score	115	487.30	441	480.658	.243	836	481.425	.26	1197	435.728	.000
Foreign	117	.068	443	.047	.363	842	.043	.215	1839	.086	.498
Home language: English	117	.880	444	.921	.165	840	.931	.052	1200	.618	.000
Home language: Afrikaans	117	.060	444	.056	.884	840	.050	.652	1200	.024	.024
Panel B: Blacks											
Female	332	.690	864	.688	.940	1774	.491	.000	3189	.502	.000
UCT admission score	326	452.34	852	449.493	.477	1750	451.995	.922	2510	451.042	.724
Foreign	332	.120	865	.074	.011	1774	.074	.004	3192	.035	.000
Home language: English	332	.569	853	.556	.672	1751	.527	.153	2511	.624	.054
Home language: Isixhosa	332	.081	853	.110	.141	1751	.112	.098	2511	.046	.005
Home language: Isizulu	332	.102	853	.095	.697	1751	.112	.612	2511	.053	.000

Notes: Data in cols. 2, 3 and 4 come from the administrative records of the university.

Table A3: Correlates of the Race IAT

Panel A: Prisoner Dilemma Game

	Full sample			*Whites*			*Blacks*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep.Var.	*Belief partner will cooperate in prisoner dilemma*								
Race IAT	0.020	0.023	-0.001	0.172	0.211	0.178	-0.082	-0.088	-0.086
	(0.043)	(0.048)	(0.050)	(0.082)	(0.088)	(0.097)	(0.054)	(0.059)	(0.061)
Index of friendship		-0.012	0.005		0.035	0.019		0.004	0.019
		(0.021)	(0.023)		(0.058)	(0.058)		(0.026)	(0.027)
Controls	--	--	X	--	--	X	--	--	X
R2	0.000	0.001	0.043	0.031	0.052	0.198	0.007	0.008	0.032
N. of Obs	508	417	417	115	92	92	342	283	283
Dep.Var.	*Cooperate in prisoner dilemma*								
Race IAT	0.018	0.061	0.042	0.051	0.155	0.129	-0.051	-0.010	-0.010
	(0.043)	(0.049)	(0.051)	(0.090)	(0.101)	(0.089)	(0.054)	(0.061)	(0.063)
Index of friendship		-0.005	0.004		0.032	0.049		0.004	0.009
		(0.022)	(0.024)		(0.060)	(0.050)		(0.028)	(0.029)
Controls	--	--	X	--	--	X	--	--	X
R2	0.000	0.004	0.062	0.003	0.029	0.342	0.003	0.000	0.038
N. of Obs	508	417	417	115	92	92	342	283	283

Panel B: Attitudes and Friendship

	Full sample		*Whites*		*Blacks*	
	(10)	(11)	(12)	(13)	(14)	(15)
Dep.Var.	*Index of attitudinal measures*					
Race IAT	0.289	0.157	0.209	0.218	0.051	0.041
	(-0.114)	(0.115)	(0.230)	(0.238)	(0.136)	(0.143)
Controls	--	X	--	X	--	X
R2	0.016	0.140	0.009	0.162	0.001	0.033
N. of Obs	467	467	107	107	310	310
	Index of friendship					
Race IAT	-0.077	0.028	-0.029	-0.093	0.032	0.007
	(0.103)	(0.097)	(0.182)	(0.206)	(0.118)	(0.118)
Controls	--	X	--	X	--	X
R2	0.001	0.225	0.000	0.062	0.000	0.070
N. of Obs	442	442	95	95	284	284

Notes: OLS Estimates with robust standard errors in parenthesis clustered at the room level. * p<.10, ** p<.05, *** p<.01. These correlations are reported at round 2 because we only ran the prisoner dilemma game in round 2. In the full sample we include the race of the respondent (white, coloured and Indian/others) with black as the omitted category and race X university residence fixed effects. In the sample of whites and blacks we include university residence fixed effects. Controls and roommate controls are measured at baseline and are described in footnote of Table 3.

Table A4: Descriptive Statistics at follow up

	Full Sample		*Mixed Room*		*Non-Mixed Room*	
	Mean	*Std. Dev.*	*Mean*	*Std. Dev.*	*Mean*	*Std. Dev.*
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full Sample (N=499)						
Race IAT	-.205	.502	-.263	.489	-.178	.507
Academic IAT	-.215	.451	-.228	.443	-.209	.455
UCT admission score	.463	.048	.466	.050	.462	.048
Wealth Index	.024	2.122	.045	1.890	.014	2.223
Consumption	.926	.847	1.021	.929	.882	.804
Foreign	.112	.316	.140	.348	.099	.300
Private high school	.601	.490	.618	.487	.594	.492
Female	.671	.470	.656	.477	.678	.468
GPA	-.011	1.005	.063	.981	-.044	1.015
Panel B: Whites (N=117)						
Race IAT	-.382	.510	-.299	.477	-.423	.523
Academic IAT	-.310	.465	-.345	.404	-.293	.495
UCT admission score	.487	.039	.487	.042	.487	.038
Wealth Index	.838	1.804	.459	1.295	1.028	1.991
Consumption	1.182	.912	1.181	1.007	1.183	.868
Foreign	.068	.253	.103	.307	.051	.222
Private high school	.744	.439	.692	.468	.769	.424
Female	.675	.470	.641	.486	.692	.465
GPA	.588	.868	.438	1.019	.663	.778
Panel C: Blacks (N=332)						
Race IAT	-.124	.487	-.214	.495	-.097	.482
Academic IAT	-.191	.441	-.210	.445	-.185	.441
UCT admission score	.453	.048	.451	.050	.453	.047
Wealth Index	-.380	2.051	-.521	1.580	-.339	2.170
Consumption	.809	.800	.894	.901	.784	.769
Foreign	.120	.326	.147	.356	.113	.317
Private high school	.536	.499	.533	.502	.537	.500
Female	.690	.463	.733	.445	.677	.469
GPA	-.234	.951	-.117	.823	-.268	.984

Notes: The difference in col. (7) is the coefficient of the dummy *Mixed Room* in a regression that includes Race X Residence fixed effects in Panel A and Residence fixed effects in Panels B and C. *UCT admission score* is the sum of high school final grades, with weights depending on the specific department the student enrolls in; *Wealth index* measures per capita ownership of durable goods in the respondent's household and is constructed applying principal component analysis to the following categories of goods: computer, fridges, TV, landline and mobile phones, bicycles, motorbikes, bakkies, electricity, gas, kettles, geysers and cars; *Consumption* is the monthly consumption (in Rands) on lunches, dinners, food, alcohol, cigarettes, cell phone minutes, entertainment; *Foreign* is a dummy equal to one if the respondent is not from South Africa; *Private high school* is equal to one if the respondent was enrolled in a private high school before joining UCT.

Table A5: Probability of being in a mixed room at baseline, simulations

Dependent variable = 1 if roommate of a different race at baseline				
	Estimates	*Simulated p-value*	*Estimates*	*Simulated p-value*
	(1)	(2)	(3)	(4)
Race IAT	-0.038	0.191	-0.021	0.319
	(0.037)		(0.039)	
Academic IAT	0.036	0.808	0.037	0.810
	(0.040)		(0.040)	
White	-0.028	0.312	-0.071	0.225
	(0.124)		(0.126)	
Coloured	0.627	0.960	0.647	0.969
	(0.234)		(0.256)	
Indian/Other	0.432	0.629	0.459	0.704
	(0.239)		(0.262)	
UCT admission score	-0.011	0.522	-0.013	0.512
	(0.446)		(0.421)	
Foreign	0.092	0.891	0.090	0.887
	(0.078)		(0.073)	
Private high school	-0.020	0.288	-0.018	0.287
	(0.042)		(0.041)	
Wealth index	-0.016	0.063	-0.014	0.091
	(0.008)		(0.008)	
Consumption	0.026	0.782	0.017	0.677
	(0.024)		(0.021)	
R-squared	0.232		0.266	
No. Obs.	499		499	

Notes: This table shows the results of a simulation exercise where we randomly assign roommates within dorms 10,000 times and compare the distribution of the simulated coefficients to the estimated coefficients of table 2. For comparison, cols. 1 and 3 report the same coefficients and standard errors as cols. 1-2 of table 2. The empirical p-values in columns 2 and 4 represent the fraction of simulations in which the simulated coefficient is smaller than the actual coefficient of cols 1 and 3, respectively.

Table A6: Placebo regression - Mixed room and lagged measures of stereotypes

Panel A:	*Dep. Var. = Race IAT (lag)*					
	Full Sample	*Whites*	*Blacks*	*Full Sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(7)	(8)	(9)
Mixed Room	-0.043	0.103	-0.083	-0.011	0.008	-0.021
	(0.055)	(0.115)	(0.070)	(0.058)	(0.128)	(0.080)
Controls	X	X	X	X	X	X
Roommate Controls	--	--	--	X	X	X
R-squared	0.088	0.078	0.048	0.129	0.157	0.092
No. Obs.	499	117	332	499	117	332
Panel B:	*Dep. Var. = Academic IAT (lag)*					
	Full Sample	*Whites*	*Blacks*	*Full Sample*	*Whites*	*Blacks*
	(4)	(5)	(6)	(10)	(11)	(12)
Mixed Room	0.045	-0.005	0.057	0.051	-0.006	0.021
	(0.055)	(0.116)	(0.066)	(0.057)	(0.127)	(0.072)
Controls	X	X	X	X	X	X
Roommate Controls	--	--	--	X	X	X
R-squared	0.071	0.177	0.065	0.099	0.245	0.107
No. Obs.	499	117	332	499	117	332
Panel C:	*Dep. Var. = Index of attitudinal measures (lag)*					
	Full Sample	*Whites*	*Blacks*	*Full Sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)
Mixed Room	0.069	0.321	-0.045	0.099	0.321	-0.045
	(0.112)	(0.288)	(0.137)	(0.112)	(0.288)	(0.137)
Controls	X	X	X	X	X	X
Roommate Controls	--	--	--	X	X	X
R-squared	0.135	0.309	0.096	0.141	0.309	0.096
No. Obs.	455	112	295	455	112	295

Notes: OLS Estimates with standard errors in parenthesis clustered at the room level. In the full sample we include the race of the respondent (White, Coloured and Indian/Others) with Black as the omitted category and Race X Residence fixed effects. In the white and black sub-samples we include Residence fixed effects. Controls and roommate controls are described in footnote of Table 3.

Table A7: Stereotypes and exposure to a roommate of different race, no roommate controls

Dependent variable:	*Race IAT*			*Academic IAT*		
Sample:	*Full Sample*	*Whites*	*Blacks*	*Full Sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)
Mixed Room	-0.029	0.210	-0.115	-0.035	0.017	-0.044
	(0.058)	(0.126)	(0.067)	(0.046)	(0.099)	(0.056)
	[0.711]	[0.201]	[0.173]	[0.711]	[0.877]	[0.432]
Controls	X	X	X	X	X	X
Mean of dep. var. in same race room	-0.178	-0.423	-0.097	-0.209	-0.293	-0.185
R-squared	0.119	0.146	0.066	0.088	0.167	0.051
No. Obs.	499	117	332	499	117	332

Notes: OLS estimates with standard errors in parenthesis clustered at the room level. Higher values of the dependent variable (IAT) indicate less prejudice against blacks. All control variables are measured at baseline. All regressions include the dependent variable at baseline. In the full sample we include the race of the respondent (White, Coloured and Indian/Others) with Black as the omitted category and Race X Residence fixed effects. In the white and black sub-samples we include Residence fixed effects. (a) Controls include IAT at baseline, a female dummy, UCT admission score, wealth index, consumption, foreign, private high school, as defined in the footnote of Table 3.

Table A8: Stereotypes and exposure to a roommate of different race, sensitivity analysis

Dependent variable:	*Race IAT*			*Academic IAT*		
Sample:	*Full Sample*	*Whites*	*Blacks*	*Full Sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)
Mixed Room	-0.005	0.493	-0.106	-0.025	-0.069	-0.007
	(0.042)	(0.094)	(0.056)	(0.038)	(0.078)	(0.049)
Controls	X	X	X	X	X	X
Roommate controls	X	X	X	X	X	X
Mean of dep. var. in same race room	-0.184	-0.436	-0.101	-0.203	-0.278	-0.184
R-squared	0.183	0.355	0.134	0.119	0.387	0.100
No. Obs.	462	105	311	468	109	311

Notes: OLS estimates on dfbeta sensitivity analysis. Standard errors in parenthesis clustered at the room level. Higher values of the dependent variable (IAT) indicate less prejudice against blacks. All control variables are measured at baseline. All regressions include the dependent variable at baseline. In the full sample we include the race of the respondent (White, Coloured and Indian/Others) with Black as the omitted category and Race X Residence fixed effects. In the white and black sub-samples we include Residence fixed effects. Controls and roommate controls are described in footnote of Table 3.

Table A9: Stereotypes and exposure to a roommate of different race, by race group

Dependent variable:	*Race IAT*			*Academic IAT*		
Sample:	*Full Sample*	*Whites*	*Blacks*	*Full Sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)
Roommate Black [A]	0.141	0.368		0.017	0.066	
	(0.063)	(0.167)		(0.060)	(0.123)	
Roommate Coloured or Indian/Asian [B]	0.064	0.189	-0.140	-0.015	-0.115	-0.019
	(0.087)	(0.172)	(0.105)	(0.074)	(0.113)	(0.085)
Roommate White [C]			-0.057			-0.000
			(0.082)			(0.077)
Controls	X	X	X	X	X	X
Roommate controls	X	X	X	X	X	X
p-value [A] = [B]	0.323	0.389		0.626	0.163	
p-value [B] = [C]			0.508			0.862
Mean of dep.var. in same race room	-0.173	-0.423	-0.096	-0.173	-0.423	-0.096
R-squared	0.135	0.223	0.099	0.105	0.273	0.087
No. Obs.	499	117	332	499	117	332

Notes: OLS estimates with standard errors in parenthesis clustered at the room level. Higher values of the dependent variable (IAT) indicate less prejudice against blacks. All control variables are measured at baseline. All regressions include the dependent variable at baseline. In the full sample we include the race of the respondent (White, Coloured and Indian/Others) with Black as the omitted category and Race X Residence fixed effects. In the white and black sub-samples we include Residence fixed effects. Controls and roommate controls are described in footnote of Table 3.

Table A10: Impact on academic performance, no roommate controls

Dependent variable:	*GPA*			*Number of exams passed*			*Eligible to continue*			*Index of Performance*		
	Full sample	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mixed Room	0.147	-0.185	0.259	0.456	-0.084	0.673	0.105	0.035	0.151	0.290	-0.082	0.449
	(0.099)	(0.259)	(0.115)	(0.193)	(0.500)	(0.225)	(0.031)	(0.074)	(0.038)	(0.109)	(0.289)	(0.130)
	[0.126]	[0.814]	[0.026]	[0.030]	[0.856]	[0.006]	[0.002]	[0.851]	[0.001]			
Controls	X	X	X	X	X	X	X	X	X	X	X	X
Mean of dep. var. in same race room	-0.044	0.663	-0.268	4.977	6.500	4.506	0.871	0.923	0.852	-0.042	0.709	-0.281
R-squared	0.412	0.542	0.369	0.702	0.689	0.709	0.307	0.330	0.372	0.433	0.371	0.430
No. Obs.	499	117	332	499	117	332	498	117	332	498	117	332

Notes: OLS estimates with standard errors in parenthesis clustered at the room level. P-values in square brackets are adjusted for multiple inference using the resampling method of Westfall and Young (1993) with 10.000 interactions. The dependent variable in cols. 1-3 is the GPA (standardized over the full sample); in cols. 4-6 it is the number of exams passed during the first year; in cols. 7-9 it is a dummy for being in good standing and eligible to continue the following year; in cols. 10-12 it is an index constructed as the first principal component of the previous three variables. Controls are measured at baseline and described in footnote of Table 3. In the full sample we include the race of the respondent (White, Coloured and Indian/Other) with Black as the omitted category and Race X Residence fixed effects; in the white and black sub-samples we include Residence fixed effects. All regressions include Academic program fixed effects.

Table A11: Impact on academic performance, sensitivity analysis

Dependent variable:	*GPA*			*Number of exams passed*			*Eligible to continue*			*Index of Performance*		
	Full sample	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mixed Room	0.052	-0.053	0.198**	0.267*	-0.156	0.522***	0.032*	0.024	0.066**	0.184**	0.111	0.350***
	(0.072)	(0.138)	(0.091)	(0.151)	(0.352)	(0.179)	(0.019)	(0.031)	(0.027)	(0.087)	(0.127)	(0.101)
Controls	X	X	X	X	X	X	X	X	X	X	X	X
Academic program FE	X	X	X	X	X	X	X	X	X	X	X	X
Mean of dep. var. in same race room	0.003	0.716	-0.182	5.064	6.574	4.598	0.921	0.956	0.900	0.074	0.757	-0.160
R-squared	0.468	0.773	0.438	0.766	0.853	0.794	0.412	0.424	0.517	0.479	0.613	0.557
No. Obs.	463	98	306	467	97	306	463	100	305	461	102	307

Notes: OLS estimates on dfbeta sensitivity analysis. Standard errors in parenthesis clustered at the room level. All regressions include Academic program fixed effects. In the full sample we include the race of the respondent (White, Coloured and Indian/Others) with Black as the omitted category and Race X Residence fixed effects. In the white and black sub-samples we include Residence fixed effects. Controls for respondent and the roommate are measured at baseline and are described in footnote of Table 3.

Table A12: Impact on academic performance in the 2nd year

Dependent variable:	*GPA*			*Number of exams passed*			*Eligible to continue*			*Index of Performance*		
Sample:	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mixed Room	-0.016	-0.082	0.052	0.805	0.428	1.012	0.089	-0.033	0.150	0.243	-0.073	0.451
	(0.052)	(0.085)	(0.070)	(0.299)	(0.808)	(0.408)	(0.039)	(0.070)	(0.062)	(0.132)	(0.252)	(0.206)
UCT admission score	3.467	7.113	1.590	6.604	-2.844	10.152	0.838	0.037	1.004	6.954	9.054	5.310
	(0.608)	(1.066)	(0.706)	(3.909)	(9.152)	(5.071)	(0.482)	(1.055)	(0.668)	(1.615)	(2.469)	(2.135)
Roommate's UCT admission score	0.739	1.318	1.090	7.255	6.750	5.793	0.013	-0.421	-0.300	2.353	2.475	2.148
	(0.408)	(0.768)	(0.548)	(2.908)	(7.272)	(3.331)	(0.296)	(0.680)	(0.414)	(1.095)	(2.234)	(1.442)
Controls	X	X	X	X	X	X	X	X	X	X	X	X
Roommate controls	X	X	X	X	X	X	X	X	X	X	X	X
Academic program fixed effects	X	X	X	X	X	X	X	X	X	X	X	X
Mean of dep. var. in same race room	0.389	0.683	0.254	5.234	6.099	4.785	0.902	0.958	0.879	-0.055	0.570	-0.349
R-squared	0.522	0.813	0.444	0.518	0.631	0.596	0.317	0.513	0.412	0.474	0.657	0.476
No. Obs.	355	105	208	355	105	208	354	105	207	354	105	207

Notes: OLS estimates with standard errors in parenthesis clustered at the room level. P-values in square brackets are adjusted for multiple inference using the resampling method of Westfall and Young (1993) with 10.000 interactions. All the regressions include Academic program fixed effects. In the full sample we include the race of the respondent (White, Coloured and Indian/Others) with Black as the omitted category and Race X Residence fixed effects. In the white and black sub-sample we include Residence fixed effects. Controls and roommate controls are measured at baseline and are described in the footnote of Table 3.

Table A13: Impact on performance and standard academic interaction, black subsample

Dep. var:	*Index of Performance*	*GPA*	*Index of Performance*	*GPA*	*Study w/ roommate*	
	(1)	(2)	(3)	(4)	(5)	(6)
Mixed Room	0.362	0.193	0.437	0.247	0.010	0.007
	(0.170)	(0.151)	(0.157)	(0.139)	(0.035)	(0.039)
Mixed Room* Same Faculty	0.187	0.124				
	(0.296)	(0.259)				
Same Faculty	0.001	0.003				0.119
	(0.175)	(0.146)				(0.048)
Mixed Room* Same Course			0.031	0.048		
			(0.289)	(0.278)		
Same Course			-0.053	0.009		
			(0.181)	(0.158)		
Respondent's UCT score	6.282	5.673	6.171	5.499	-0.054	-0.138
	(1.766)	(1.500)	(1.621)	(1.393)	(0.410)	(0.422)
Roommate's UCT score	1.168	0.967	1.322	1.090	-0.173	-0.207
	(1.365)	(1.249)	(1.358)	(1.241)	(0.312)	(0.317)
Controls	X	X	X	X	X	X
Mean of dep. var. in same race room	-0.285	-0.275	-0.281	-0.268	0.066	0.069
R-squared	0.453	0.380	0.447	0.386	0.160	0.193
No. Obs.	308	308	325	325	328	308

Notes: OLS estimates with standard errors in parenthesis clustered at the room level. The variable "Same course" indicates at least one course in common between respondent and the roommate.

Table A14: Impact on friendships

Dependent variable:	*No. Times hang out with people of different race over past month*			*Last time hang out with people of different race*			*% of friends of a different race (excl. roommate)*		
Sample:	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mixed Room	0.698	1.608	0.566	-0.774	-1.423	-0.631	0.111	0.143	0.063
	(0.252)	(0.594)	(0.292)	(0.283)	(0.853)	(0.331)	(0.032)	(0.057)	(0.040)
	[0.028]	[0.043]	[0.256]	[0.028]	[0.265]	[0.256]	[0.005]	[0.079]	[0.393]
Controls	X	X	X	X	X	X	X	X	X
Roommate controls	X	X	X	X	X	X	X	X	X
	3.705	4.216	3.524	1.832	1.392	1.984	0.153	0.130	0.157
R-squared							0.264	0.329	0.142
No. Obs.	481	111	320	480	111	319	462	110	306
Dependent Variable	*% of study-mates of a different race*			*Preferred number of people of different race in:*					
				Leisure group			*Academic group*		
Sample:	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Mixed Room	0.060	0.151	0.045	0.251	0.403	0.125	0.244	0.134	0.297
	(0.040)	(0.091)	(0.045)	(0.165)	(0.224)	(0.226)	(0.152)	(0.250)	(0.194)
	[0.276]	[0.265]	[0.537]	[0.276]	[0.265]	[0.574]	[0.276]	[0.594]	[0.393]
Controls	X	X	X	X	X	X	X	X	X
Roommate controls	X	X	X	X	X	X	X	X	X
Mean of dep. var. in same race room	0.171	0.148	0.163	2.720	2.320	2.788	2.849	2.547	2.908
R-squared	0.277	0.433	0.095	0.168	0.299	0.082	0.148	0.265	0.100
No. Obs.	438	97	296	483	112	322	483	112	321

Notes: Cols. 1-6 report ordered logit estimates; cols. 7-18 OLS estimates with standard errors in parenthesis clustered at the room level. P-values in square brackets are adjusted for multiple inference using the resampling method of Westfall and Young (1993) with 10.000 interactions. In the full sample we include the race of the respondent (White, Coloured and Indian/Others) with Black as the omitted category and Race X Residence fixed effects. In the white and black sub-samples we include Residence fixed effects. Controls and roommate controls are measured at baseline and are described in footnote of Table 3. *No. times hang out more with people of different race in the last month*: =0 if never. =1 if once. =2 if 2-5 times. =3 if 5-10 times. =4 if more than 10 times. *Last time hang out with people of different race*: =0 if never. =1 if last year. =2 if last month. =3 if last week. =4 if yesterday.

Table A15: Impact on explicit attitudes

Dependent variable:	*Talked about race*			*Comfortable talking about race*			*Disagree to abolish affirmative action*			*Not conscious dancing with a person of another race*			*Not conscious having boyfriend/girlfriend of another race*		
Sample:	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Mixed Room	0.430	-0.756	1.017	0.056	0.016	0.043	-0.011	0.003	0.010	0.088	0.193	-0.001	0.106	0.191	0.041
	(0.226)	(0.575)	(0.277)	(0.042)	(0.090)	(0.053)	(0.051)	(0.129)	(0.053)	(0.042)	(0.084)	(0.055)	(0.049)	(0.135)	(0.059)
	[0.168]	[0.540]	[0.003]	[0.345]	[0.981]	[0.872]	[0.821]	[0.982]	[0.977]	[0.163]	[0.148]	[0.980]	[0.154]	[0.540]	[0.872]
Controls	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Roommate controls	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Mean of dep. var. in same race room	1.550	1.472	1.564	0.799	0.829	0.797	0.740	0.423	0.835	0.775	0.671	0.808	0.667	0.347	0.765
R-squared				0.097	0.215	0.121	0.235	0.152	0.080	0.139	0.423	0.106	0.268	0.426	0.124
Observations	456	108	300	445	106	292	451	107	296	449	106	296	453	108	297

Notes: Cols. 1-3 report ordered logit estimates; cols. 4-15 OLS estimates. Standard errors in parenthesis clustered at the room level. P-values in square brackets are adjusted for multiple inference using the resampling method of Westfall and Young (1993) with 10.000 interactions. In the full sample we include the race of the respondent (White, Coloured and Indian/Others) with Black as the omitted category and Race X Residence fixed effects. In the white and black sub-samples we include Residence fixed effects. All controls are measured at baseline and are described in footnote of Table 3. Talked about race (ordinal): "In the last month. how often did you talk with any friends of yours about topics of discrimination and racial bias?" 1 Never. 2 Rarely. 3 Sometimes. 4 Most of the times. 5 Always. Comfortable talking about race: =1 if comfortable talking to people about race. Disagree to abolish Affirmative Action: = 1 if does not agree that that affirmative action in University admission should be abolished. Not conscious dancing with a person of another race: = 1 if respondent does not feel conscious dancing with a person of another race. Not conscious having a boyfriend/girlfriend of another race = 1 if respondent does not feel conscious having a boyfriend/girlfriend of another race.

Table A16: Impact on pro-social behavior

Dependent variable:	*Member of Volunteer Organization*			*Money given to a charity*			*Cooperate in Prisoner dilemma*			*Belief partner will cooperate in prisoner dilemma*		
Sample:	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mixed Room	0.100	0.203	0.047	64.155	113.494	32.840	0.082	0.259	0.095	0.073	0.140	0.096
	(0.062)	(0.113)	(0.073)	(76.134)	(300.461)	(42.021)	(0.057)	(0.108)	(0.069)	(0.055)	(0.115)	(0.069)
	[0.405]	[0.229]	[0.697]	[0.448]	[0.739]	[0.697]	[0.423]	[0.019]	[0.523]	[0.423]	[0.227]	[0.523]
Controls	X	X	X	X	X	X	X	X	X	X	X	X
Roommate controls	X	X	X	X	X	X	X	X	X	X	X	X
Mean of dep. var. in same race room	0.449	0.446	0.455	130.162	252.101	81.393	0.555	0.500	0.570	0.605	0.592	0.605
R-squared	0.084	0.339	0.063	0.166	0.256	0.129	0.117	0.359	0.061	0.076	0.265	0.045
No. Obs.	467	110	312	405	99	260	493	114	330	493	114	330

Notes: OLS estimates with standard errors in parenthesis clustered at the room level. P-values in square brackets are adjusted for multiple inference using the resampling method of Westfall and Young (1993) with 10.000 interactions. In the full sample we include the race of the respondent (White, Coloured and Indian/Others) with Black as the omitted category and Race X Residence fixed effects. In the white and black sub-samples we include Residence fixed effects. Controls and roommate controls are measured at baseline and are described in footnote of Table 3. In cols 7-12. controls also include a dummy indicating whether the respondent knows the partner in the game.

Table A17: Impact on friendships, attitudinal measures and pro-social behaviour, no missing values

Dependent variable:	*Index of friendship*			*Index of attitudinal measures*			*Index of pro-social behavior*			*Global Index of social behavior*		
Sample:	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*	*Full sample*	*Whites*	*Blacks*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mixed Room	0.444	0.447	0.371	0.238	0.455	0.076	0.168	0.427	0.220	0.541	0.655	0.380
	(0.116)	(0.182)	(0.155)	(0.102)	(0.210)	(0.131)	(0.119)	(0.229)	(0.139)	(0.128)	(0.227)	(0.165)
Controls	X	X	X	X	X	X	X	X	X	X	X	X
Roommate controls	X	X	X	X	X	X	X	X	X	X	X	X
R-squared	0.298	0.400	0.135	0.161	0.322	0.069	0.113	0.272	0.046	0.251	0.354	0.133
No Obs.	499	117	332	499	117	332	499	117	332	499	117	332

Notes: OLS estimates with standard errors in parenthesis clustered at the room level. The indices used as dependent variables are constructed using a polychoric principal component analysis. The index of friendship (cols.1-3) includes the following variables: (i) no. of times respondent hung out with people of different race in the last month: =0 if never. =1 if once. =2 if 2-5 times. =3 if 5-10 times. =4 if more than 10 times; (ii) last time respondent hung out with people of different race: =0 if never. =1 if last year. =2 if last month. =3 if last week. =4 if yesterday; (iii) fraction of friends of a different race (excl. roommate); (iv) preferred number of people of different race in leisure group and academic group. The index of attitudes (cols.4-6) includes the following variables: (i) "In the last month. how often did you talk with any friends of yours about topics of discrimination. prejudice and racial bias?": =0 if never. =1 if rarely. =2 if sometimes. =3 if most of the time. =4 if always; (ii) a dummy for whether respondent is comfortable talking about race; (iii) a dummy for whether respondent does not agree that affirmative action in University admission should be abolished; (iv) a dummy for whether respondent do not feel conscious dancing with a person of another race; (v) dummy for whether respondent do not feel conscious having a boyfriend/girlfriend of another race. The index of pro-social behavior (cols.7-9) includes the following variables: (i) member of community service or volunteer organization; (iii) dummy for whether respondent cooperated in the prisoner dilemma game; (iii) dummy for whether respondent believed partner would cooperate in prisoner dilemma. The global index of social behavior (cols.10-12) includes all the variables listed in the previous three indexes. In the full sample we include the race of the respondent (White, Coloured and Indian/Others) with Black as the omitted category and Race X Residence fixed effects. In the white and black sub-sample we include Residence fixed effects. Controls and roommate controls are described in footnote of Table 3.