

---

Interaction and Implicit Bias: heterogeneity by  
student characteristics

Abe Keijts (576087)

---



---

Supervisor:	Teresa Marreiros Bago d'Uva
Second assessor:	Philip Hans Franses
Date final version:	1st July 2023

---

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam. This paper replicates the core part of Corno et al. [2022] and extents it. Part of the replication results are added to the Appendix.

## Abstract

Interracial contact is a commonly studied way of reducing inter-group conflict. In this study, I consider students exogenously allocated a roommate in residences of the University of Cape Town. I find that being assigned a roommate of another race significantly reduces White students' implicit bias towards Black individuals. This effect of interracial contact seems to be slightly strengthened by the amount of implicit bias towards the Black race by the roommate. The effect seems to be well-balanced across own implicit bias and roommate academic ability. Besides reducing implicit bias, interracial contact in this setting also tends to improve actual interracial contact and behaviour for White students.

## 1 Introduction

This paper investigates the impact of interracial contact on interracial conflict. I make use of a policy in place at the University of Cape Town (UCT) in South Africa, which is meant to generate interracial contact. It randomly assigns students to university residences and, in some cases, also randomly assigns roommates. As students are exogenously allocated to rooms, this allows me to establish a causal relation between interracial contact and measures of prejudice. Exogenous allocation tackles the possibility of reverse causality when roommate allocation is, for example, determined by the students' preferences, or other possibly relevant characteristics. A better understanding of the effects of interracial contact allows policymakers to improve the effectiveness of policies designed to alleviate interracial conflict. This is especially relevant with the many introductions of policies with these aims, such as affirmative action on college campuses.

In this study, I address the following main research-question: "Does engaging with individuals of a different race alter an individual's stereotypes concerning that race?" I use data from students residing in UCT residences who were exogenously allocated to roommates. The sample consists of 499 freshmen residing in double rooms. The students were interviewed both at the beginning and the end of the 2012 academic year. The primary focus was on examining prejudices or stereotypes held against individuals from different racial groups. One of the measures used are implicit association tests (IATs). The IAT was initially introduced by Greenwald and Banaji [1995] and serves as a measurement of implicit associations, which is commonly used in the field of psychology. Two different IATs were conducted at both rounds. The first being what is referred to as the Race IAT, which is intended to capture general positive or negative associations with race. The second is an IAT designed to capture implicit bias regarding academic ability and race, which I call the Academic IAT. The advantage of utilizing IATs over self-reported measures of prejudice is that it prevents from a possible reporting bias. Considering the tense atmosphere around race issues, certainly in South-Africa, this feature may be especially useful for this kind of research. While there is skepticism regarding the predictive power and reliability of IAT scores within the field of psychology ([Blanton et al., 2009], [Oswald et al., 2013]), the IAT is commonly utilized in social psychology as a measure of implicit bias. I aim to establish the size of the effect of being in a mixed room on one's implicit attitudes, as measured by the IATs. Additionally, I investigate the existence of heterogeneity of the effect of interracial contact by certain characteristics, which is more thoroughly discussed in Section 3. The primary

finding of this study is that exposure to individuals of a different race has a significant impact on stereotypes. Specifically, I observed a notable decrease in prejudice among White students towards Black students. This effect is substantial, corresponding to approximately 0.63 standard deviations of the Race IAT. This estimated effect indicates that the treatment would effectively reduce the disparity in stereotypes between White students and Black students, as measured by the Race IAT. However, I did not observe such changes in the Academic IAT. This may be the result of this measure being more a reflection of statistical knowledge about the actual differences in academic performances between Black and White students. I do not find evidence of heterogeneity of the effect of being in a mixed room on the Race IAT by respondents' own baseline IAT, indicating that the effect may be well balanced among more and less prejudiced individuals. This may be an encouraging result, as it suggests that interracial contact may not only benefit those who already have low levels of prejudices, but also those with stronger negative views about the other race. However, one should be cautious about this result, as the relatively small sample may make it difficult to detect such heterogeneity. The IAT score of one's roommate generally does not significantly explain one's own IAT score at a post-exposure measurement. However, there may be a small effect for White students in mixed race rooms. Controlling for one's roommate implicit bias yields no significant results while being in a mixed room maintains a significantly positive effect, meaning the estimated effect of interracial contact is likely not caused by peer imitation effects. For roommate high school performance, I find some remarkable results. There seems to be a non-linear effect of roommate performance on implicit prejudice for White students paired with another White student. This effect, however, cancels out for White students in mixed race rooms. Similar results have not been found in the recent literature, and I cannot think of a plausible explanation for this seemingly remarkable relationship.

Additionally, I make use of a number of self-reported attitudinal measures and behavioral outcomes. Findings for those measures indicate that exposure to a roommate of a different race leads to an increase in interracial interactions outside the room as well. Students in mixed rooms reported spending more time socializing with individuals from different racial backgrounds. Moreover, they expressed a higher desire and had a greater number of actual friends and study partners from different races compared to students in same-race rooms. Additionally, self-reported attitudes showed improvement among students in mixed rooms.

This paper shows evidence of the positive effects of interracial contact on multiple outcomes. Most importantly, it shows a positive effect on implicit attitudes, something studied only very limited in the existing literature. Secondly, it seems that certain behaviours and explicit attitudes may be drivers of this effect, although no causal effect was established. Thirdly, this effect seems to be well-balanced along a set of different characteristics, although I do find some evidence of heterogeneity of the effect of being in a mixed room by roommate implicit bias for White students. Interestingly, I find a non-linear relation between roommate academic ability and own IAT score for White students paired with another White student, while I find no such effect for White students in mixed race rooms.

The rest of the paper is structured as follows. Section 2 provides a background of the existing literature related to this research. In Section 3, I explain the sub-questions used to answer the

main research-question. Section 4 gives an explanation about the origin and characteristics of the data. Subsequently, in Section 5 I will give the regression models used to answer the sub-questions and motivate the use of them. Section 6 provides the results, and finally, Section 7 provides a conclusion.

## 2 Literature

This paper relates to literature on the effect of interracial contact on attitudes and behaviours, which is heavily influenced by Allport [1954]. Allport’s “contact hypothesis” states that contact between members of groups with different characteristics could lead to a reduction of prejudices and improve inter-group relations, given favorable circumstances. Those circumstances include the following prerequisites: (1) individuals exposed to out-group members should have equal status, (2) these individuals have common goals, (3) cooperation should be beneficial for both groups, and (4) authorities involved endorse interaction. A commonly studied setting in which those conditions can be argued to apply are educational institutions. Besides Allport’s conditions, colleges sometimes offer another useful instrument to estimate the effect of inter-group contact on attitudes and behaviour: random roommate allocation. Exogenously determined intensities of inter-group contact allows to better establish a causal effect. Van Laar et al. [2005] found that among American college students, those randomly allocated to a roommate of another ethnic group showed reduced prejudices toward other races, which is in line with the contact hypothesis. However, they found that the group the roommate belongs to matters. Most interestingly, they found that among White and Black individuals, those allocated to an Asian American roommate actually showed signs of worsened ethnic prejudices towards other races, and in the case of Black individuals, sometimes even towards their own race. The authors give as possible explanation that Asian Americans had significantly higher base-line prejudices than other ethnic groups. They suggest that roommates of Asian Americans tended to modify their attitudes and behaviors to be more in line with those of their peers [Feldman and Newcomb, 1969]. Boisjoly et al. [2006] also considered randomly allocated students. They found that White students who are randomly assigned African American roommates are significantly more likely to be in favour of affirmative action policies and have personal contact with members of other ethnic groups after their first year. Carrell et al. [2019] investigate male freshmen among the US Air Force Academy, who were randomly allocated into squadrons. They found that White males exposed to Blacks in the top two terciles of high school performance exhibit increased future interactions with Blacks. However, they do not find such effect for exposure to the lower tercile performing Blacks. The authors therefore conclude that affirmative action policies that involve lowering admission standards for minorities may be ineffective at increasing interaction with minorities among majorities. Other notable studies investigating inter-group contact in seemingly favourable circumstances include Scacco and Warren [2018], which studies explicit attitudes and outcomes in behavioral games for a high conflict area. Results were less optimistic in this setting, as no significant reductions in prejudice were found for inter-group contact. Mousa [2020] also investigates the effect on behavioral outcomes and explicit attitudes, for Iraq Christians and Muslims. While a positive effect was found for the setting within the intervention, there was not a considerable amount of spillover to everyday life.

The above mentioned studies all suggest there might indeed be an inter-group conflict reducing effect of inter-group contact under favourable circumstances, as proposed by Allport [1954]. However, results were less optimistic for high-conflict areas. Especially remarkable findings are those of Van Laar et al. [2005] and Carrell et al. [2019], suggesting that the effect of interracial contact on prejudices towards other races may be effected by third variables, namely the amount of bias towards other races of a student’s roommate and the high school performance of Black peers. Those two findings provide the argumentation for two branches of sub-questions in this paper, as will be explained in Section 3. While the aforementioned studies focus on self-reported attitudinal measures or behavioral observations, Corno et al. [2022] complements this by adding measures for implicit attitudes towards the other race. More specifically, they use IAT’s to examine the effects of interaction between randomly allocated roommates at the University of Capetown.

## **3 Theory**

### **3.1 First sub-question**

Not only is it important to reduce prejudices among those with mild prejudices, more importantly, societies should aim to reduce prejudices among those with the strongest negative views regarding the other race. As discussed in Section 2, many studies indicate the positive effects of interracial contact on prejudices. I aim to establish whether this effect is evenly distributed among students, or whether this effect is heterogeneous by one’s baseline prejudice level. As such, I pose the following sub-question: “To what extent is the effect of being allocated a mixed race room heterogeneous by baseline prejudice level?”

### **3.2 Second set of sub-questions**

Van Laar et al. [2005] found that interracial contact with Asian Americans tended to increase prejudice towards other races among White students, and could even decrease the positive affect Blacks felt towards other Blacks. They give as possible explanation for this seemingly contradictory finding to Allport [1954] that Asian Americans were significantly more prejudiced than the other ethnic groups involved. As such, they suggest that this finding is likely caused by peer imitation [Feldman and Newcomb, 1969]. Or, more specifically, when students are exposed to more prejudiced individuals, they tend to become more prejudiced themselves, as they adjust their behaviors to better match that of their peers. As such, my second sub-question is the following: “What is the effect of roommates’ implicit racial attitudes on an individual’s own implicit racial attitudes?” An additional question to follow up on this is: “Is this effect different for exposure to other race roommates than for same race roommates?” Lastly, to better assess whether contact between two individuals with different levels of prejudices improves the overall level of prejudice in a population, I pose the following sub-question: “Are the effects of roommate IAT of the same magnitude for the most and least prejudiced individual in a shared room?” A greater negative effect of being in a shared room with someone more prejudiced than the positive effect the other way around may not be helpful in reducing the aggregate amount of prejudice found in a population.

### 3.3 Third sub-question and motivation

Carrell et al. [2019] found that on average, when White males are exposed to Black students from the top two terciles of high school performance, it increases the willingness to pair up with a Black roommate later in their educational period. However, they find no positive nor negative effect of exposure to Black students from the bottom third of the high school performance distribution. In order to investigate whether academic performance has a similar effect in this experiment, I consider the following sub-question: “What is the moderating effect of out-group roommates’ academic ability on the influence of interracial contact on a White student’s own implicit racial attitudes?”

## 4 Data

### 4.1 Setting and design of experiment

As briefly mentioned in Section 1, the data originates from an experiment from the UCT. The UCT is a renowned university that admits approximately 5,000 freshmen per year. There are two main criteria UCT uses to determine acceptance of students. Firstly, they utilize the Admission Points Score (APS), which is a measure calculated based on high school grades during the final year. Secondly, there is a policy in place that aims to ensure a diverse ethnic student population is admitted. In 2012, the incoming freshmen class comprised 36% Black, 16% Coloured, 40% White, and 8% Indian, Asian, or other race. When students live on campus, they are randomly assigned one of the 15 residences, and either a one-person or two-person room. When assigned a two-person room, a student is always paired with someone of the same gender. Each residence has a warden who is in charge of the room allocation. What criteria they use differ among the wardens. I focus on eight residences, of which the wardens stated they had used a random allocation. It is important to note that room assignments are not permanent. First-year students have the option to request room or residence changes. Approximately 19% of the students interviewed at the follow-up reported changing roommates since the beginning of the academic year in the residences of interest. Only two of those students moved from a mixed room to a non-mixed room, and none the other way around. For my analysis, I will use the initial room assignments. The data was collected in two different rounds. The first being the baseline round, conducted at the start of the academic year. The follow-up data collection was held at the end of the academic year. A series of IATs were administered in both rounds. Additionally, during the follow-up survey, lab experiments were conducted with the same individuals who had participated in the baseline survey. In order to obtain estimates of explicit attitudes, additional questionnaires were conducted as well. In order to attract respondents, the project was mentioned in a meeting with wardens and students, it was advertised using posters, and emails were sent to all students in the eight residences of interest. The subject of research was always said to be “students life at UCT”. The word “race” was not mentioned on purpose, to prevent the subjects from being informed about the real goal of the research. This aimed to prevent reporting bias among the respondents. Additionally, the IAT scores were conducted first in each round, to reduce reporting bias as well. Initially, the sample consisted of 621 freshmen who completed the survey and had non-missing IAT scores at the baseline. Out of these, 499

participated in the follow-up. In Appendix A, I analyze the factors influencing whether the students participated in the follow-up round. The follow-up rate was not significantly influenced by being in either a mixed or non-mixed room, measures for stereotypes and interactions between stereotype measures and being in a mixed room. It seems that women and rich students tended to have a lower chance of participating in the follow-up round. The effect of wealth may be explained by the low pay-off of participation, which was only 3.50 dollars.

Appendix B draws several comparisons. The sample of interest is compared to four other samples. As shown, the sample of interest is in most ways very similar to other students, however, they tend to have higher admission scores and are more likely to be female than students not living in campus residences or students from other years.

## 4.2 IATs

Implicit attitudes are measured using IAT scores, which are commonly used in psychology. The IAT was first introduced by Greenwald and Banaji [1995]. The aim of IATs is to detect subconscious associations between concepts. Participants are instructed to pair representations of certain attributes to concepts. A shorter response time is seen as an indication of a stronger association, while a longer response time is seen as an indication of a more difficult pairing, and thus a weaker association. The advantage that IAT scores have over explicit attitude measures, such as questions in surveys, is that it likely prevents answers from being manipulated to be more socially desirable. Considering the tense atmosphere around race issues, certainly in South-Africa, this feature may be especially useful for this kind of research. I use two distinct types of IAT measures. The first is a standard test which had participants pair positive and negative attributes, such as “good” and “terrible” with the concepts of White or Black South Africans. This test, in line with existing literature (Blanton et al. [2009]; Oswald et al. [2013]), is referred to as the Race IAT and captures taste-based associations. The second IAT is a test custom-designed to capture associations between academic ability and race. Participants were instructed to match pictures of individuals from different races with percentiles representing high and low performance in the UCT grading system. Examples of used categories include: “first class (> 75%)” and “third class (< 60%)”. The Academic IAT primarily measures beliefs regarding academic performance instead of prejudice. Additional details on the procedures of both IATs can be found in the Online Appendix of Corno et al. [2022].

In Figure 1, one can observe the densities of both IAT scores for the White and Black sub-sample. Both tests are designed in such a way that lower values correspond to more negative stereotypes toward Blacks than to Whites. As can be seen, the Race IAT distribution of the White sub-sample is more concentrated for lower values than the Black sub-sample. This indicates that White respondents tended to have more negative prejudices towards Blacks than Black respondents, as one might expect. The mean of the Race IAT was also negative for the Black sub-sample<sup>1</sup>. When looking at the Academic IAT score distributions, I can see little difference for the White and Black sub-sample. Additionally, both sub-samples had a negative mean. This could indicate that the Academic IAT is more a reflection of the objective knowledge about av-

---

<sup>1</sup>Implicit bias towards one’s own group is commonly found in the existing literature. For example, Rudman and Kilianski [2000] found negative implicit bias by female respondents towards female authority.

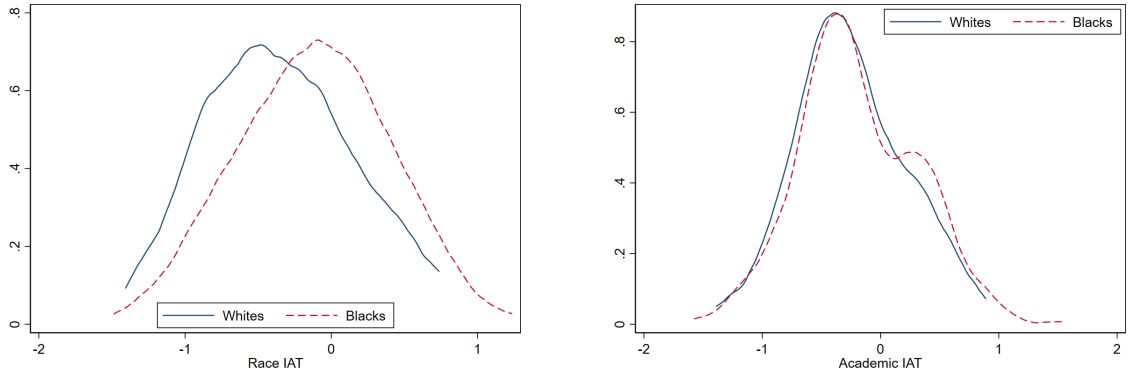


Figure 1: IAT score densities at baseline

erage academic performances of White and Black students, rather than a subjective taste-based measure.

### 4.3 IATs validity and limitations

Despite the aforementioned advantages of IATs, criticism exists on the validity and limitations of IATs, as some authors question whether implicit attitudes are actually a good indicator of discriminatory behavior. In a meta-analysis, Greenwald et al. [2009] argue that many studies have found that implicit attitudes can actually be a good indicator of actual discriminatory behaviour. However, Oswald et al. [2013] criticise that they overestimated the positive correlation between IAT scores and behaviour, by including studies that did not directly examine actual discriminatory behaviour. Blanton et al. [2009] also question the predictive power of IATs regarding discriminatory behaviour. For the sample in this study, I find some evidence IATs are useful for predicting behaviour. Results for this can be found in Appendix C. Secondly, IATs may overestimate bias for respondents with less difference in reaction times, as the differences are divided by standard deviations in the scoring algorithm. This critique does not play a major role for the purpose I use IATs for, as I am interested in the difference between individuals in mixed and non-mixed rooms, and I control for baseline IAT scores. It is unlikely that one group would become less consistent in their responses than the other group. Additionally, an IAT score that may seem prejudiced could be due to negative associations with a certain race, that do not necessarily have to do with discriminatory views ([Arkes and Tetlock, 2004], [Uhlmann et al., 2006]). For example, many people are aware that minorities, due to their disadvantaged societal position tend to be less academically skilled. This does not have to mean however, that a person automatically thinks they are less talented.

Another critique is that IAT scores can be highly sensitive to short term factors. Dasgupta and Greenwald [2001] have shown that IAT scores tend to become less biased when respondents are exposed to pictures of Black and White Americans with positive and negative reputations, respectively. However, as I essentially compare change in IAT score for treatment and non-treatment groups, this issue may not be very problematic, as there is no clear reason why one group would be more or less influenced by such short term factors. As a last critique, Fiedler and Bluemke [2005] have shown that respondents instructed to fake an IAT are generally able



to do so. The IAT's were the first test score conducted in this experiment, to mitigate the threat of adjusting one's responses to be more socially acceptable. Moreover, IAT tests were not commonly used in South Africa during the conduction of this experiment. An important caveat here, that seems to be overlooked in Corno et al. [2022], is that students are in fact familiar with the IAT test at follow-up. This may cancel out the intention of the set-up of the experiment, and bias the results towards a greater reduction in implicit attitudes. However, as I am mostly interested in the differences between students in mixed and non-mixed rooms, this does not necessarily need to be a problem, as long as this learning effect is similar across both groups. One could argue that the bias could have a higher effect for the non-mixed room group, as the treated group's responses are already expected to be closer to what is considered socially desirable. In that case, this would deflate the effect of the treatment. It seems unlikely that learning effects would inflate the effect of the treatment. It may thus be that this biases me to not rejecting the hypothesis of no effect of being in a mixed room, which I actually do reject. In short, the IAT has some potential drawbacks. Due to the purpose the IAT scores are used for and the setting of the experiment, most of those concerns can be argued to have little effect.

#### **4.4 Explicit attitudes and behaviour**

To capture explicit attitudes, I use results of survey questionnaires. The questions were on three main topics: friendship patterns, interracial attitudes and prosocial behaviour. The specific questions can be found in Appendix E. Additionally, for the follow-up round, an experimental measure of prosocial behaviour was collected by means of a prisoner's game. For the experimental instructions, I refer to Corno et al. [2022]. Finally, through the administrative data at UCT, I obtained information on residence and room choices for the sample for the second year. Using this information, three binary variables were constructed. The first equals one when the respondent is still in the same residence in year 2. The second equals one when the respondent is still in the same residence and in a mixed room in the second year. Lastly, the third equals one when the respondent is still in the same residence and has the same roommate in year 2.

#### **4.5 Descriptive Statistics and Randomization Check**

In total, the dataset includes 499 students, who had non-missing values at both baseline and follow-up. Of this sample, 332 student were Black, 117 were White, 18 were coloured and 32 belonged to a group from Indian, Asian or other descent. In total, 157 students were allocated a room with someone of another race category, while 342 were paired with someone of their own race group. This means that 31% of students were allocated a mixed room. To compare the observed pairings in the actual sample to a truly random allocation, a series of resampling-type tests were conducted. A simulation with 10,000 replications results in an average probability of 0.52 for a student to end up in a mixed room. This is significantly higher than the previously mentioned probability of 0.31 for the observed sample. This may indicate that the wardens in the residences of interest did not apply a truly random room allocation. Even though I find this significant difference between the observed sample and the findings obtained from the simulation, additional tests show that room allocation may still be considered as good as random for the specific purpose of this research. In Table 1, one can observe descriptive statistics for the full

sample and different sub-samples at baseline. The sub-samples considered are the mixed room and non-mixed room sub-samples for the full sample as well as the sub-samples of White and Black students (columns 3,4,5 and 6). Columns 1 and 2 give descriptives for the full sample and the full White and Black sub-samples (so here, mixed room and non-mixed room are taken together, whereas they are split for columns 3,4,5 and 6). Included in the table are the outcomes of interest: Race IAT and Academic IAT, as well as the following control variables: UCT entry score, wealth, consumption at baseline, a dummy variable taking value one when a student is foreign, and a dummy variable taking value one in case the student attended a private high school. Column 7 gives the estimate of the coefficient of the dummy variable Mixed Room in a regression that includes race x residence effects for the full sample and residence effects for the sub-samples. Column 8 gives the P-value on the hypothesis that this coefficient is equal to zero. As can be seen from column 8, none of the variables are significantly different by the variable Mixed Room. Not only are the outcome variables similar for the mixed and non-mixed groups overall, but also when only considering the White and Black sub-samples. The control variables also seem to be well-balanced across the sub-samples. In Table 2, I present results of a regression check. The dependent variable is always a binary variable taking value one when a respondent was allocated a mixed race room, and zero else. The explanatory variables include the key outcome measures, as well as control variables. Columns 1 and 2 show results for the full sample for regressions with different controls included. The same can be seen for the White and Black sub-samples, in columns 3, 4, 5 and 6. The key outcome variables Race IAT and Academic IAT are insignificant for all models and samples. The dummy variable for Coloured is significant at the 5% level in the full sample for both regressions. The dummy for Indian/other is significant for the first regression only at the 10% level. Those findings are not surprising, as individuals from small groups are more likely to be matched with someone of another group. In Table 3, one can observe the results of a regression of the same binary outcome as in Table 2, with as explanatory variables an index of explicit attitudes at baseline and different sets of controls. The table includes regressions for the full sample and the White and Black sub-samples. Again, being in a mixed room or not does not significantly depend on explicit attitudinal measures. As an additional check, I have performed a dyadic regression. This considers all possible combinations of roommates and considers the differences in characteristics of individuals  $i$  and  $j$  as explanatory variables for whether or not those individuals were allocated the same room. The approach taken is similar to Caprettini [2020] and Fafchamps and Gubert [2007]. The model used is the following:

$$SameRoom_{ij} = \alpha + \beta|X_i - X_j| + \gamma(X_i + X_j) + \epsilon_{ij} \quad (1)$$

Here,  $SameRoom_{ij}$  is a binary variable taking value one when student  $i$  and  $j$  are in the same room.  $X$  denotes a vector of explanatory variables. It includes baseline measurements of the IAT scores, an index of attitudinal measures, UCT entry score, as well as a set of demographic characteristics. As can be observed in Table 4, I do not find any of the differences in IAT scores or explicit attitudes to have a significant effect on being in the same room. UCT entry score and consumption both show significant effects. However, none of the other demographic variables considered are significant.

Table 1: Descriptive statistics for the full sample and White and Black sub-samples

(a) Descriptive statistics full sample

Variable	Full sample		Mixed rooms		Not mixed rooms		Beta	Beta=0
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Diff	P-value
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 entry score	0,463***	0,048	0,466***	0,050	0,462***	0,048	0,000	0,923
Wealth	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
1 if foreign	0,112	0,316	0,140	0,348	0,099	0,300	0,037	0,286
1 if private school	0,601	0,490	0,618	0,487	0,594	0,492	-0,021	0,688
Observations	499		157		342		499	

(b) Descriptive statistics White sub-sample

Variable	Full Sample		Mixed Rooms		Not Mixed Rooms		Beta	Beta=0
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Diff	P-value
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 entry score	0.487***	0.039	0.487***	0.042	0.487***	0.038	0.003	0.725
Wealth	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
1 if foreign	0.068	0.253	0.103	0.307	0.051	0.222	0.067	0.210
1 if private school	0.744	0.439	0.692	0.468	0.769	0.424	-0.079	0.390
Observations	117		39		78		117	

(c) Descriptive statistics Black sub-sample

Variable	Full Sample		Mixed Rooms		Not Mixed Rooms		Beta	Beta=0
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Diff	P-value
Race IAT	-0.119	0.520	-0.174	0.538	-0.103	0.514	-0.081	0.240
Academic IAT	-0.188	0.498	-0.139	0.514	-0.203	0.494	0.062	0.341
UCT entry score	0.453***	0.048	0.451***	0.050	0.453***	0.047	-0.001	0.833
Wealth	-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 school	0.536	0.499	0.533	0.502	0.537	0.500	0.001	0.983
Observations	332		75		257		332	

Notes: UCT entry score is the weighted average of high school scores in the final years, divided by 1,000. Wealth is constructed using Principal Component Analysis using the following goods: computer, fridges, TV, landline, mobile phones, bicycles, motorbikes, pick-up trucks, electricity, gas, kettles, geysers, and cars. Consumption represents the expenditure per month on: lunches, dinners, food, alcohol, cigarettes, cell phone minutes, and entertainment. Foreign is a dummy variable equal to one when the student is not from South Africa. Private school is a dummy variable taking value one when the student attended a private high school. The *Betadiff* values in Column 7 are the coefficients of the dummy variable *MixedRoom* in a regression that includes race x residence effects in Panel (a) and residence fixed effects in Panels (b) and (c).

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

	Full sample		Whites		Blacks	
	1	2	3	4	5	6
Race IAT	-0.038 (0.037)	-0.021 (0.039)	0.075 (0.084)	-0.002 (0.082)	-0.068 (0.047)	-0.023 (0.049)
Academic IAT	0.036 (0.040)	0.037 (0.040)	-0.019 (0.117)	-0.016 (0.113)	0.057 (0.048)	0.020 (0.048)
White	-0.028 (0.124)	-0.071 (0.126)				
Coloured	0.627*** (0.234)	0.647** (0.256)				
Indian/Other	0.432* (0.239)	0.459* (0.262)				
UCT entry score	-0.011 (0.446)	-0.013 (0.421)	0.494 (1.214)	0.858 (1.254)	-0.089 (0.520)	-0.106 (0.494)
1 if foreign	0.092 (0.078)	0.090 (0.073)	0.199 (0.198)	0.118 (0.196)	0.083 (0.090)	0.058 (0.084)
1 if private school	-0.020 (0.042)	-0.018 (0.041)	-0.161 (0.112)	-0.141 (0.120)	0.001 (0.050)	0.003 (0.049)
Wealth	-0.016** (0.008)	-0.014* (0.008)	-0.031 (0.021)	-0.041* (0.024)	-0.013 (0.010)	-0.007 (0.010)
Consumption	0.026 (0.024)	0.017 (0.021)	0.030 (0.055)	0.039 (0.050)	0.028 (0.029)	0.021 (0.025)
Roommate controls		X		X		X
Residence effects			X	X	X	X
Race x residence effects	X	X				
$R^2$	0.232	0.266	0.214	0.367	0.037	0.143
Number of observations	499		117		332	

Notes: Estimates are obtained using OLS. Standard errors in parentheses are clustered at the room level.

Table 3: Probability of being in a mixed room by explicit attitudes

	1	2	3	4	5	6
Attitudes index	0.011 (0.018)	0.014 (0.017)	0.050 (0.037)	0.042 (0.040)	-0.004 (0.025)	-0.007 (0.023)
Mean of dep. var.	0.000	0.000	0.000	0.000	0.000	0.000
$R^2$	0.238	0.268	0.241	0.390	0.031	0.137
Number of observations	455	455	112	112	295	295
Roommate controls	X		X		X	
Residence fixed effects			X	X	X	X
Race x residence fixed effects	X	X				

Notes: Estimates are obtained using OLS. Standard errors in parentheses are clustered at the room level. The Attitudes index is constructed using PCA.

Table 4: Dyadic regression

Variable	Coefficient	SE
Absolute difference in Race IAT for (ij)	0,000905	(0.001)
Absolute difference in Academic IAT for (ij)	0.000576	(0.001)
Absolute difference in Attitude index for (ij)	0.000307	(0.000)
Absolute difference in UCT entry score	-0.000950***	(0.000)
Absolute difference in Private school for (ij)	-0.000794	(0.001)
Absolute difference in Wealth for (ij)	-0.000030	(0.000)
Absolute difference in Consumption for (ij)	-0.001330**	(0.001)
Absolute difference in Foreign for (ij)	-0.002940	(0.002)
Number of observations		59,522

Notes: Standard errors in parentheses are corrected for dyadic correlation following Fafchamps and Gubert [2007] and Caprettini [2020].

## 5 Methodology

### 5.1 IAT models

#### 5.1.1 Model for effect of being in mixed room

Firstly, I replicate the model as used by Corno et al. [2022], with the aim of answering the question whether interracial contact reduces prejudices in this study’s setting. The model is specified as follows:

$$IAT_{ijk1} = \beta_1 IAT_{ijk0} + \beta_2 MixedRoom_{ik0} + \gamma_1 X_{ik0} + \gamma_2 X_{jk0} + \gamma_3 \delta_k + \epsilon_{ijk1} \quad (2)$$

Here,  $IAT_{ijk1}$  is the IAT score for individual  $i$  paired with individual  $j$ , in residence  $k$ , at time  $t = 1$ , which indicates the IAT was measured at the end of the first academic year. The regressions will be performed for both IAT scores.  $IAT_{ijk0}$  is the same variable for time  $t = 0$ , which represents the baseline measurement.  $MixedRoom_{ik0}$  is a dummy variable indicating whether individuals  $i$  and  $j$  were allocated into a mixed race room at time  $t = 0$ .  $X_{ik0}$  and  $X_{jk0}$  are vectors including control variables for individual  $i$  and  $j$ , respectively. Those controls include gender, UCT entry score, household wealth, the student’s monthly consumption expenditure, and a dummy variable indicating whether the respondent is not from South Africa.  $\delta_k$  includes seven residence dummies. Finally,  $\epsilon_{ijk1}$  represents the error term. The model is estimated for the Black and White sub-samples separately.

#### 5.1.2 Heterogeneity by own IAT score

In order to assess whether the effect of being in a mixed room is heterogeneous by one’s pre-exposure level of prejudice towards the Black race, I propose the following model:

$$IAT_{ijk1} = \beta_1 IAT_{ijk0} + \beta_2 MixedRoom_{ik0} + \beta_3 (MixedRoom_{ik0} * IAT_{ijk0}) + \beta_4 IAT_{jik0} + \gamma_1 X_{ik0} + \gamma_2 X_{jk0} + \gamma_3 \delta_k + \epsilon_{ijk1} \quad (3)$$

This model contains the interaction term  $MixedRoom_{ik0} * IAT_{ijk0}$  with corresponding coefficient  $\beta_3$ . This coefficient gives the heterogeneity of the effect of being in a mixed room by one’s own baseline IAT score.

### 5.1.3 Heterogeneity by roommate's IAT score

Subsequently, I want to address the question whether the effect of being in a mixed room is heterogeneous by one's roommates' implicit attitudes. In order to do this, I perform the following moderation analysis:

$$IAT_{ijk1} = \beta_1 IAT_{ijk0} + \beta_2 MixedRoom_{ik0} + \beta_3 (MixedRoom_{ik0} * IAT_{jik0}) + \beta_4 IAT_{jik0} + \gamma_1 X_{ik0} + \gamma_2 X_{jk0} + \gamma_3 \delta_k + \epsilon_{ijk1} \quad (4)$$

This model contains the interaction term ( $MixedRoom_{ik0} * IAT_{jik0}$ ) and the control variable  $IAT_{jik0}$ . My coefficients of interest are  $\beta_2, \beta_3$  and  $\beta_4$ .  $\beta_4$  gives the effect of roommate IAT for non-mixed rooms, which is taken as the baseline level.  $\beta_3$  gives the difference in effect of being in a mixed room by one's roommate's IAT score.

Additionally, I want to investigate whether the effect of roommate IAT is symmetric. More specifically, I want to establish whether the most prejudiced individual is more strongly effected by roommate IAT than the least prejudiced individual, or vice versa. In order to do this, I propose the following model:

$$IAT_{ijk1} = \beta_1 IAT_{ijk0} + \beta_2 MixedRoom_{ik0} + \beta_3 (PREJ_{ik0} * IAT_{jik0}) + \beta_4 IAT_{jik0} + \gamma_1 X_{ik0} + \gamma_2 X_{jk0} + \gamma_3 \delta_k + \epsilon_{ijk1} \quad (5)$$

Here,  $PREJ_{ik0}$  represents a dummy indicating whether person  $i$  was the room member with the lowest IAT score (so the most prejudiced toward Black individuals). As such,  $\beta_3$  gives the difference between the effect of roommate IAT for the most and least prejudiced individual. This allows me to test whether the effect of IAT is symmetric.

### 5.1.4 Heterogeneity by roommate academic ability

To address my third sub-question, I propose the following model:

$$IAT_{ijk1} = \beta_1 IAT_{ijk0} + \beta_2 MixedRoom_{ik0} + \beta_3 (MixedRoom_{ik0} * UCT_j) + \gamma_1 X_{ik0} + \gamma_2 X_{jk0} + \gamma_3 \delta_k + \epsilon_{ijk1} \quad (6)$$

This model contains the interaction  $MixedRoom_{ik0} * UCT_j$ .  $UCT_j$  represents the UCT entry score for individual  $j$ , standardized to have mean zero and standard deviation equal to one to make interpretation easier. My main coefficient of interest here is  $\beta_3$ , as this gives the heterogeneity of the effect of being assigned a mixed room by one's roommate's high school performance.

To address the possible non-linear nature of this interaction effect, I consider the following model

$$IAT_{ijk1} = \beta_1 IAT_{ijk0} + \beta_2 MixedRoom_{ik0} + \beta_3 (MixedRoom_{ik0} * UCT_j) + \beta_4 (MixedRoom_{ik0} * UCT_j^2) + \gamma_1 X_{ik0} + \gamma_2 X_{jk0} + \gamma_3 \delta_k + \epsilon_{ijk1} \quad (7)$$

Here, I have added the same interaction term as in the previous model, and additionally added its square. This model is capable of catching a smooth non-linear relation between the interaction term and the dependent variable.

## 5.2 Explicit attitudes and behaviours

In order to estimate the effect of being allocated a mixed race room on explicit and behavioural outcomes, the following model is used:

$$Y_{ijk1} = \beta_1 \text{MixedRoom}_{ik0} + \gamma_1 X_{ik0} + \gamma_2 X_{jk0} + \gamma_3 \delta_k + \epsilon_{ijk1} \quad (8)$$

Here,  $Y_{ijk1}$  denotes the outcome variable of interest. For attitudes, these are the index of friendships, index of explicit attitudes, index of prosocial behaviour and a global index of social behaviour. The indices are described more thoroughly in Appendix E. The behavioural outcomes include three binary variables indicating residential choices in the second year, as explained in Section 4.4. The control variables are the same as in previous equations. Note that no lagged value of the outcome measure is included, as those are not measured at baseline. Note that I did not perform a moderation analysis using explicit outcomes or behavioural outcomes, as only one interaction effect turns out to be significant at the 10% level, as I will show in Section 6. Additionally, while the meaning of the IATs are on prejudices towards the Black race, the explicit and behavioural measures have different meanings (see Appendix E for the exact questions).

## 6 Results

### 6.1 IATs

#### 6.1.1 Effect of being in a mixed room

Table 5 shows the estimation results of Equation (2). The results are reported for the White and Black sub-sample separately. Standard errors in parentheses are clustered at room-level. Being allocated a mixed race room has a significantly positive effect on the Race IAT for the White sub-sample. This indicates that White students show a significant reduction in implicit prejudices when paired with someone of another race. For the Black sub-sample, the effect is not significantly different from zero. Interesting is the notable size of the effect for the White sub-sample: 0.316, which amounts to 0.63 standard deviation of the Race IAT. Such effect puts the average White student at the same level of prejudice as the average Black student, when paired with someone of another race. As for the Academic IAT, being paired with a roommate of a different race has no significant effect for both sub-samples. This indicates that the Academic IAT is more a reflection of the differences in academic performances of White and Black students one finds in reality, than a measure of prejudice.

#### 6.1.2 Heterogeneity by baseline IAT score

Table 6 displays the estimation results for Equation (3). Note that this and subsequent regression models are only carried out for the Race IAT. As shown by the earlier findings, being allocated a mixed race room does not significantly alter one's beliefs on academic performances of Blacks relative to Whites. The coefficient of the interaction between one's own baseline Race IAT and being in a mixed room is insignificant at the 10% level for the White and Black sub-samples. This suggests the positive effect of being allocated a mixed race room seems to be well balanced between individuals with different pre-exposure prejudice levels. Note that the relatively small sample used makes it more difficult to detect such heterogeneity.

Table 5: Effect of being in a mixed race room on IAT scores

Dep. Variable sample	RaceIAT		Academic IAT	
	Whites	Blacks	Whites	Blacks
Mixed Room	0.316** (0.140) [0.047]	-0.094 (0.069) [0.361]	0.014 (0.107) [0.904]	-0.009 (0.059) [0.889]
Controls	X	X	X	X
Roommate Controls	X	X	X	X
Mean of dep. var.	-0.423	-0.097	-0.293	-0.185
$R^2$	0.217	0.097	0.266	0.087
Number of observations	117	332	117	332

Notes: Standard errors in parentheses are clustered at the room-level. Here, P-values are adjusted for multiple hypothesis testing using the re-sampling method of Westfall and Young [1993].

Table 6: Heterogeneity of the effect of being in a mixed room by baseline IAT

Dep. Variable sample	RaceIAT	
	Whites	Blacks
Mixed Room	0.299 (0.136) [0.030]	-0.107 (0.073) [0.145]
Baseline Race IAT	0.106 (0.132) [0.423]	0.067 (0.065) [0.303]
Baseline Race IAT*Mixed Room	-0.057 (0.239) [0.814]	-0.077 (0.129) [0.554]
Controls	X	X
Roommate Controls	X	X
Mean of dep. var.	-0.423	-0.097
$R^2$	0.2179	0.0984
Number of observations	117	332

Notes: Standard errors in parentheses are clustered at the room-level.



### 6.1.3 Heterogeneity by roommate's baseline IAT score

Columns 1 and 2 of Table 7 show the estimation results for Equation (4). As described in Section 4.3, I have added roommate's IAT as control variable, and an interaction effect between roommate IAT and the variable Mixed Room. Note that the amount of observations are slightly lower for each sub-sample, as the regression only includes respondents for whom the IAT scores of their room partners were recorded as well. This made the size of the White sub-sample decrease from 117 to 75, and the size of the Black sub-sample from 332 to 215. As can be seen, for the White sub-sample the interaction effect is positive and significant only at the 10% level. This may indicate there is a slight strengthening effect of exposure to another race, when the roommate in question has more positive attitudes towards the Black race. The effect of roommate IAT is not significant at the 10% level. The effect of being in a mixed room is still positive and significant at the 5% level. For the Black sub-sample, Mixed Room remains insignificant at the 10% level. The added variable Roommate IAT and the interaction effect are both insignificant at the 10% level as well.

Table 7: Heterogeneity of the effect of being in a mixed room by roommate's baseline IAT

Model sample	Interaction		Symmetry	
	Whites	Blacks	Whites	Blacks
Mixed Room	0.417 (0.193) [0.036]	-0.165 (0.109) [0.132]	0.275 (0.177) [0.126]	-0.139 (0.091) [0.130]
Roommate IAT	-0.151 (0.189) [0.429]	-0.110 (0.076) [0.149]	-0.157 (0.240) [0.517]	-0.154 (0.096) [0.111]
Roommate IAT*Mixed Room	0.547 (0.303) [0.078]	-0.077 (0.188) [0.684]		
Roommate IAT*roommate higher IAT			0.495 (0.392) [0.214]	0.083 (0.166) [0.620]
Controls	X	X	X	X
Roommate Controls	X	X	X	X
Mean of dep. var.	-0.390	-0.075	-0.390	-0.075
$R^2$	0.235	0.0948	0.2287	0.0951
Number of observations	75	215	75	215

Notes: Columns 1 and 2 show results for the model including an interaction effect between Mixed Room and Roommate IAT. Columns 3 and 4 show results of the symmetry check of the effect of Roommate IAT. Standard Errors in parentheses are clustered at the room-level again, and P-values are given in square brackets.

Columns 3 and 4 of Table 7 show results of the symmetry check as described in equation (5). For the White sub-sample, roommate IAT does not have an effect significantly different from zero, at the 10% level. As the interaction effect is also not significantly different from zero at the 10% level, this indicates I cannot find an effect of roommate IAT for both White students that are the least and most prejudiced individual in the room. Note that in this regression, the effect of Mixed Room is no longer significant at the 10% level either. This may be due to the increased standard error (0.140 vs. 0.177), which in turn may be caused by the smaller sample used, as described in the beginning of this subsection. For the Black sub-sample, again no significant

effect of any of the variables of interest was found.

#### 6.1.4 Heterogeneity by roommate academic ability

In Table 8, I show the results of the estimations of Equation (6) and Equation (7). In the first two columns, I show the results for the linear model for the White and Black sub-sample. For the White sub-sample, the coefficient of the interaction between Roommate UCT entry score and Mixed Room is insignificant at the 10% level. This suggests I cannot find a similar sort of heterogeneity by out-group academic ability as was found by Carrell et al. [2019]. For the Black sub-sample, Mixed Room has no significant effect, which is in line with the findings of the earlier models. The interaction between Mixed Room and Roommate UCT entry score is also insignificant for the Black sub-sample. In Columns 3 and 4, I show the results for the

Table 8: Heterogeneity of the effect of being in a mixed room by roommate academic ability

Model sample	Linear		Non-linear	
	Whites	Blacks	Whites	Blacks
Mixed Room	0.352 (0.125) [0.006]	-0.095 (0.069) [0.171]	0.236 (0.142) [0.101]	-0.037 (0.079) [0.645]
Roommate UCT entry score	0.107 (0.090) [0.239]	0.020 (0.033) [0.537]	0.372 (0.104) [0.001]	0.039 (0.039) [0.320]
<i>Roommate UCT entry score</i> <sup>2</sup>			-0.231 (0.099) [0.022]	0.018 (0.019) [0.337]
Roommate UCT entry score*Mixed Room	-0.114 (0.142) [0.421]	0.020 (0.070) [0.773]	-0.350 (0.173) [0.046]	-0.022 (0.078) [0.782]
<i>Roommate UCT entry score</i> <sup>2</sup> * Mixed Room			0.273 (0.144) [0.061]	-0.058 (0.037) [0.123]
Controls	X	X	X	X
Roommate Controls	X	X	X	X
Mean of dep. var.	-0.4234	-0.0972	-0.4234	-0.0972
<i>R</i> <sup>2</sup>	0.2248	0.0976	0.2719	0.0976
Number of observations	117	332	117	332

Notes: Columns 1 and 2 show estimates for the model only including a linear interaction, while Columns 3 and 4 also include a quadratic interaction. Standard errors in parentheses are clustered at the room-level again, and P-values are given in square brackets.

non-linear model including a squared variable *Roommate UCT entry score*<sup>2</sup> and a squared interaction *Roommate UCT entry score*<sup>2</sup> \* *Mixed Room*. For the White sub-sample, all interactions and the *Roommate UCT entry score* and *Roommate UCT entry score*<sup>2</sup> are individually significant. Performing a joint test on the significance of *Roommate UCT entry score* and *Roommate UCT entry score*<sup>2</sup> gives a P-value of 0.0022, indicating there indeed is an effect of roommate UCT entry score in case a students is not allocated a mixed room. Performing a joint test on the significance of *Roommate UCT entry score* \* *Mixed Room* and *Roommate UCT entry score*<sup>2</sup> \* *Mixed Room* gives a P-value of 0.0193, indicating there also is an interaction effect. From this output, I can write the IAT scores for Whites in non-mixed

rooms approximately as:

$$IAT_{ijk1} = 0.372 * UCT_j - 0.231 * UCT_j^2 + \text{terms without Roommate UCT entry score} \quad (9)$$

As such, the turning point is at a Roommate UCT entry score of approximately 0.8 standard deviations above the mean. This corresponds to an original UCT entry score of approximately 0.5. Note that the transformed UCT entry scores are on a scale from 0 to 0.6. Below 0.8 standard deviations above the mean, increases in Roommate UCT entry score increase the Race IAT, thus making individuals more prejudiced. Values of Roommate UCT entry score greater than 0.8 standard deviations above the mean decrease the Race IAT, thus making students less prejudiced. I performed tests to establish whether the sums of the non-interacted and interacted coefficients were equal to zero, and they both did not reject the sum to be equal to zero (P-values 0.8802 and 0.6817 for the linear and quadratic variables, respectively). Overall, there seems to be a non-linear effect of roommate UCT score for White students in non-mixed rooms, while for White students in mixed rooms, I find no significant effect of roommate UCT, as the interacted coefficients cancel out the non-interacted coefficients.

## 6.2 Attitudinal and behavioural measures

Table 9: Index of Friendship, Explicit Attitudes, Prosocial Behaviour, Global Social Behaviour

	White			
	Friendship Index	Explicit Attitudes Index	Prosocial Behaviour	Global Social Behaviour
Mixed Room	0.477** (0.187) [0.044]	0.670** (0.261) [0.044]	0.438* (0.250) [0.088]	0.760** (0.294)
Mean of dep. var.	-1.053	-1.643	-0.873	-1.604
R <sup>2</sup>	0.505	0.369	0.374	0.458
Number of observations	94	106	94	79
	Black			
Mixed Room	0.254 (0.170) [0.368]	0.072 (0.166) [0.664]	0.229 (0.165) [0.368]	0.196 (0.212)
Mean of dep. var.	-1.173	-0.565	-0.651	-1.465
R <sup>2</sup>	0.155	0.068	0.099	0.149
Number of observations	275	299	253	203
	Full			
Mixed Room	0.340*** (0.128) [0.030]	0.318** (0.126) [0.030]	0.169 (0.138) [0.230]	0.439*** (0.150)
Mean of dep. var.	-1.110	-0.813	-0.705	-1.457
R <sup>2</sup>	0.317	0.186	0.168	0.321
Number of observations	411	453	388	315
Controls	X	X	X	X
Roommate controls	X	X	X	X

Notes: The panels give estimates for White and Black sub-samples, and the full sample of the effect of Mixed Room on the indices. Standard errors in parentheses are clustered at the room level again, and P-values are given in square brackets.

In Table 9, I show the results for the regressions with attitudinal and behavioral measures. Through a survey, 15 different measures were collected. Those measures are grouped into three

main categories, which are used as outcome variables for the regressions. For an overview of which measure was allocated to which index, see Appendix E. Additionally, I consider a global index, which is constructed using PCA on all 15 measures. As can be read from the first column of Table 9, being in a mixed room significantly increases the friendship index for White students. For Blacks, the effect does not differ significantly from zero. In the full sample, the effect is significantly positive. The results for explicit attitudes are very similar to those of friendships. The effect is again insignificant for the Black sub-sample and significantly positive for the White sub-sample and full sample. For the index of prosocial behaviour, the effect is significant at the 10% level for White students only, and insignificant for the Black sub-sample and full sample. For the global index, the results are similar to those in the first two columns. Again, the effect is significantly positive for the full sample, which is mainly driven by the significantly positive effect found in the White sub-sample. The effect in the Black sub-sample was again not significantly different from zero.

Overall, one can conclude that the pattern seen in explicit attitudes seems to match the results found for implicit attitudes, as strong positive effects are found for the White sub-sample.

To further test behaviour changes, I consider residence choices for the same sample in the second year. As can be seen in the first panel of Table 16 in Appendix F, being allocated a mixed race room in the first year does not significantly alter the probability of being in the same residence in the second year, for both sub-samples considered and the full sample. Similar results are found in the second panel, indicating that the decision to remain in the same residence and being in a mixed room is not significantly influenced by being allocated a mixed room in the first year. Finally, in the last panel, one can observe that being allocated a mixed room in the first year does not significantly correlate with being in the same residence and having the same room partner in the second year. In short, it seems that being allocated a mixed room in the first year does not alter behaviour regarding room decisions in the second year.

## 7 Conclusions

Exposure to members of different groups in positive settings is often seen as beneficial for reducing inter-group conflict. While many authors have studied the effect of inter-group contact on explicit measures of prejudice, this study aimed to establish whether such positive effects can also be found for implicit attitudes, using IAT scores. To this end, I posed the following research question: “Does engaging with individuals of a different race alter an individual’s stereotypes concerning that race?” In order to answer this research-question, I used data originating from an experiment conducted at the University of Cape Town, which had students exogenously allocated to either mixed race or same race double rooms. The results show a decrease of implicit prejudices towards Black individuals among White students paired with a roommate of another race. This effect seems to be well-balanced along a set of different characteristics, with only weak evidence of heterogeneity by roommates’ implicit attitudes. Interestingly, I find a non-linear relation between implicit prejudices and roommate high school performance for White students paired with White students. More specifically, for roommate UCT entry scores lower than 0.5, an increase in roommate score tends to go along with an increase in the Race IAT at follow-up, which corresponds to a decrease in bias against Black individuals. For

roommate scores higher than 0.5, an increase in roommate score generally corresponds to a decrease in Race IAT at follow-up. Note that a UCT entry score of 0.5 is relatively high, as it corresponds to approximately 0.8 standard deviations above the mean. To the best of my knowledge, no such effects have yet been found in the existing literature. I do not manage to find a plausible explanation for the nature of this effect, and why it is cancelled out for White students in mixed race rooms. I am unsure whether a similar effect of peer academic ability can be found in slightly different settings, or if this effect is specifically for this sample. Future research may be needed to gain a better understanding of the role of peer academic ability on racial attitudes. Additionally, I show positive effects of being in a mixed room for a multitude of behavioural outcomes and explicit attitudes. Corno et al. [2022] state these may be seen as potential drivers of the positive effect of interracial contact on implicit attitudes, although no causal effect on these potential drivers of implicit attitudes can be inferred from the performed analyses. It remains for future research to establish whether these measures are indeed drivers of the prejudice reducing effect of interracial contact.

Next, there are some important limitations to this paper to discuss. First of all, the relatively small size of the sample used may have limited the statistical power to find evidence of heterogeneity of the effect of interracial contact. This may be even more a limiting factor for the models including a variable representing the implicit attitudes of a respondent's roommate, as this forced me to leave out observations that had a roommate that had not been tracked in the experiment. Secondly, one should be cautious in using the results of this study as an argument to justify the use of affirmative action policies. Not only can it be argued that affirmative action is a form of discrimination itself, it also remains unclear how, for example, other students react to such policies when they read about it in the media, as was also brought up by Carrell et al. [2019]. An obvious example of a group for whom affirmative action policies may be argued to have adverse effects are those that belong to the "non-oppressed" group that may be negatively affected. In the setting of this experiment, that could be White students not admitted to UCT, who have scores that would have allowed them to enroll had they been Black. It remains for further research to investigate the effect of affirmative action among groups affected in different ways by affirmative action policies. In conclusion, the main takeaway for society and policymakers is the large potential of inter-group contact in reducing interracial conflict.

## References

- G. W. Allport. *The nature of prejudice*. ADDISON-WILEY PUBL, 1954.
- H. R. Arkes and P. E. Tetlock. Target article: Attributions of implicit prejudice, or "would jesse jackson 'fail' the implicit association test?". *Psychological Inquiry*, 15(4):257–278, 2004. doi: 10.1207/s15327965pli1504\_01.
- H. Blanton, J. Jaccard, J. Klick, B. Mellers, G. Mitchell, and P. E. Tetlock. Strong claims and weak evidence: Reassessing the predictive validity of the iat. *Journal of Applied Psychology*, 94(3):567–582, 2009. doi: 10.1037/a0014665.
- J. Boisjoly, G. J. Duncan, M. Kremer, D. M. Levy, and J. Eccles. Empathy or antipathy? the

- impact of diversity. *American Economic Review*, 96(5):1890–1905, 2006. doi: 10.1257/aer.96.5.1890.
- B. Caprettini. Team visibility and city travel: Evidence from the uefa champions’ league random draw. *Journal of Sports Economics*, 22(1):85–114, 2020. doi: 10.1177/1527002520955208.
- S. E. Carrell, M. Hoekstra, and J. E. West. The impact of college diversity on behavior toward minorities. *American Economic Journal: Economic Policy*, 11(4):159–182, 2019. doi: 10.1257/pol.20170069.
- L. Corno, E. La Ferrara, and J. Burns. Interaction, stereotypes, and performance: Evidence from south africa. *American Economic Review*, 112(12):3848–3875, 2022. doi: 10.1257/aer.20181805.
- N. Dasgupta and A. G. Greenwald. 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, 2001. doi: 10.1037/0022-3514.81.5.800.
- M. Fafchamps and F. Gubert. The formation of risk sharing networks. *Journal of Development Economics*, 83(2):326–350, 2007. doi: 10.1016/j.jdeveco.2006.05.005.
- K. A. Feldman and T. M. Newcomb. *The impact of college on students*. Transaction Publishers, 1969.
- K. Fiedler and M. Bluemke. Faking the iat: Aided and unaided response control on the implicit association tests. *Basic and Applied Social Psychology*, 27(4):307–316, 2005. doi: 10.1207/s15324834basp2704\_3.
- A. G. Greenwald and M. R. Banaji. Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1):4–27, 1995. doi: 10.1037/0033-295x.102.1.4.
- A. G. Greenwald, T. A. Poehlman, E. L. Uhlmann, and M. R. Banaji. Understanding and using the implicit association test: Iii. meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1):17–41, 2009. doi: 10.1037/a0015575.
- S. Mousa. Building social cohesion between christians and muslims through soccer in post-isis iraq. *Science*, 369(6505):866–870, 2020. doi: 10.1126/science.abb3153.
- F. L. Oswald, G. Mitchell, H. Blanton, J. Jaccard, and P. E. Tetlock. Predicting ethnic and racial discrimination: A meta-analysis of iat criterion studies. *Journal of Personality and Social Psychology*, 105(2):171–192, 2013. doi: 10.1037/a0032734.
- L. A. Rudman and S. E. Kilianski. Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin*, 26(11):1315–1328, 2000. doi: 10.1177/0146167200263001.
- A. Scacco and S. S. Warren. Can social contact reduce prejudice and discrimination? evidence from a field experiment in nigeria. *American Political Science Review*, 112(3):654–677, 2018. doi: 10.1017/s0003055418000151.

- E. L. Uhlmann, V. L. Brescoll, and E. L. Paluck. 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, 2006. doi: 10.1016/j.jesp.2004.10.003.
- C. V. Van Laar, S. Levin, S. Sinclair, and J. Sidanius. The effect of university roommate contact on ethnic attitudes and behavior. *Journal of Experimental Social Psychology*, 41(4):329–345, 2005. doi: 10.1016/j.jesp.2004.08.002.
- P. H. Westfall and S. S. Young. *Resampling-based multiple testing: Examples and methods for P-Value Adjustment*. Wiley, 1993.

## A Attrition check

Table 10 shows estimates for models used to check correlates of attrition. As can be seen, attrition only seems to be somewhat higher for females and wealthier students. The last result may be explained by the fact that respondents received a relatively low compensation (approximately 3.50 dollars) for participation.

## B Sample comparison

Table 11 shows comparisons of the sample used in this research to the following samples: Students in the 8 residences (single and double rooms) not in the sample, students in all residences (single and double rooms) not in the sample, all first year students (residence and non residence) not in the sample, and all students in all years (residence and non residence) not in the sample. As can be seen, the sample used does not differ much in the shown characteristics from other students in the same residences. However, compared to students not living on campus and from other years, students in the used sample tend to have higher UCT entry scores and are more likely to be female.

## C Predictive power of IAT scores

Table 12 shows regression results of two binary outcome variables. In the first panel, the dependent variable takes value one when respondents believe the partner will cooperate in the prisoner’s dilemma game, held at the follow-up round. As can be seen, for White students, a higher Race IAT score (and thus less negative prejudices towards the Black race) tends to increase the chance the corresponding student believes the partner will cooperate. This even continues to hold when controlling for stated racial preferences, as can be seen in Column 3. This suggests the Race IAT can add information on top of self-reported explicit attitudes. In the second panel, the dependent variable takes on value one when the respondent decides to cooperate in the prisoner’s game. As can be seen, no significant effects are found for this outcome. In Table 14, a similar check is shown for the index of attitudinal measures and the index of friendships. No significant effect of the Race IAT was found for either outcome variable, for any of the sub-samples and models used.

Table 10: Correlates of attrition

Variable	1	2	3	4	5	6	7
Mixed Room	-0.009 (0.043)		-0.007 (0.044)	0.004 (0.044)	-0.009 (0.043)	-0.011 (0.043)	
White	-0.474* (0.249)	-0.470* (0.252)	-0.491* (0.252)	-0.467* (0.251)	-0.522** (0.258)	-0.450* (0.269)	-0.410 (0.270)
Coloured	-0.123 (0.144)	-0.128 (0.147)	-0.135 (0.149)	-0.119 (0.146)	-0.125 (0.152)	-0.171 (0.157)	-0.041 (0.171)
Indian/Other	0.043 (0.117)	0.041 (0.116)	0.031 (0.117)	0.048 (0.119)	0.038 (0.125)	0.030 (0.165)	0.060 (0.121)
Female	-0.138** (0.054)	-0.140*** (0.054)	-0.139*** (0.053)	-0.137** (0.055)	-0.137** (0.055)	-0.137** (0.055)	-0.136** (0.055)
UCT entry score	0.332 (0.378)	0.333 (0.379)	0.351 (0.381)	0.334 (0.380)	0.348 (0.382)	0.291 (0.381)	0.335 (0.383)
1 if foreign	0.091 (0.057)	0.090 (0.056)	0.090 (0.057)	0.090 (0.057)	0.095* (0.057)	0.095* (0.057)	0.094* (0.057)
1 if private school	-0.026 (0.035)	-0.026 (0.036)	-0.026 (0.036)	-0.027 (0.036)	-0.026 (0.036)	-0.028 (0.036)	-0.027 (0.036)
Wealth	-0.003 (0.009)	-0.003 (0.009)	-0.003 (0.009)	-0.003 (0.009)	-0.003 (0.009)	-0.003 (0.009)	-0.004 (0.009)
Consumption	-0.046* (0.024)	-0.046* (0.024)	-0.046* (0.024)	-0.046* (0.024)	-0.046* (0.024)	-0.046* (0.024)	-0.046* (0.024)
Race IAT		0.005 (0.031)		0.002 (0.036)		0.003 (0.036)	
Academic IAT			0.023 (0.031)		0.002 (0.036)		0.022 (0.035)
Race IAT x Mixed Room					0.010 (0.066)		
Aca. IAT x Mixed Room						0.066 (0.065)	
White x Race IAT							0.047 (0.076)
Coloured x Race IAT							-0.201 (0.185)
Other x Race IAT							-0.022 (0.209)
White x Academic IAT							0.005 (0.075)
Coloured x Academic IAT							0.104 (0.156)
Other x Academic IAT							-0.090 (0.108)
Mean of dep. var.	0.804	0.804	0.804	0.804	0.804	0.804	0.804
$R^2$	0.126	0.126	0.126	0.126	0.128	0.129	0.128

Notes: Standard Errors are clustered at room-level. Controls include race x residence fixed effects.



Table 11: Sample Comparisons

	the Sample			Students in the 8 residences			Students in all residences			All first year students			All students in all years		
	N	Mean	P-value	N	Mean	P-value	N	Mean	P-value	N	Mean	P-value	N	Mean	P-value
Panel B: Black															
Female	332	0.690	0.688	864	0.688	0.940	1774	0.491	0.000***	3191	0.502	0.000***	9919	0.533	0.000***
UCT entry score	326	452.344	852	449.493	0.477	451.995	1750	451.995	0.922	2510	451.027	0.721	6862	454.241	0.62
Foreign	332	0.120	0.074	865	0.074	0.011**	1774	0.074	0.004***	3194	0.035	0***	9740	0.041	0***
Home language: English	332	0.569	0.556	853	0.556	0.672	1751	0.527	0.153	2512	0.623	0.057*	6863	0.645	0.005***
Home language: Isizulu	332	0.081	0.110	853	0.110	0.141	1751	0.112	0.098*	2512	0.046	0.005***	6863	0.029	0***
Home language: Isixhosa	332	0.102	0.095	853	0.095	0.697	1751	0.112	0.612	2512	0.053	0***	6863	0.029	0***
Home language: Afrikaans	332	0.015	0.005	853	0.005	0.065*	1751	0.005	0.026**	2512	0.031	0.097*	6863	0.031	0.097*
Panel C: White															
Female	117	0.675	0.707	444	0.707	0.502	844	0.495	0***	1845	0.517	0.001***	4909	0.512	0***
UCT entry score	115	487.330	441	480.658	0.243	481.425	836	481.425	0.26	1199	435.692	0***	3482	435.730	0***
Foreign	117	0.068	0.047	443	0.047	0.363	842	0.043	0.215	1842	0.086	0.501	4886	0.116	0.111
Home language: English	117	0.880	0.921	444	0.921	0.165	840	0.931	0.052*	1202	0.618	0***	3486	0.542	0***
Home language: Isizulu	117	0.000	0.000	444	0.000	.	840	0.002	0.598	1202	0.027	0.07*	3486	0.057	0.008***
Home language: Isixhosa	117	0.000	0.000	444	0.000	.	840	0.002	0.598	1202	0.042	0.025**	3486	0.056	0.009***
Home language: Afrikaans	117	0.060	0.056	444	0.056	0.884	840	0.050	0.652	1202	0.024	0.024**	3486	0.015	0***

Notes: P-values are on the test whether the mean of the corresponding variable and sub-sample is significantly different from the same variable for the sample used in this study.

Table 12: Relation Race IAT and behaviour

	Whites		Blacks			
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.: Belief partner will cooperate in prisoner dilemma						
Race IAT	0.172** (0.082)	0.211** (0.088)	0.178* (0.097)	-0.082 (0.054)	-0.088 (0.059)	-0.086 (0.061)
Friendship index		0.035 (0.058)	0.019 (0.058)		0.004 (0.026)	0.019 (0.027)
Controls			X			X
$R^2$	0.031	0.052	0.198	0.007	0.008	0.032
Number of observations	115	92	92	342	283	283
Dep. Var.: Decision to cooperate in prisoner's dilemma						
Race IAT	0.051 (0.090)	0.155 (0.101)	0.129 (0.089)	-0.051 (0.054)	-0.010 (0.061)	-0.010 (0.063)
Friendship index		0.032 (0.060)	0.049 (0.050)		0.004 (0.028)	0.009 (0.029)
Controls			X			X
$R^2$	0.003	0.029	0.342	0.003	0.000	0.038
Number of observations	115	92	92	342	283	283

Notes: Standard Errors are clustered at the room-level.

Table 13: Relation Race IAT and explicit attitudes

	Whites		Blacks	
	(1)	(2)	(3)	(4)
Dep. Var.: Index of attitudinal measures				
Race IAT	0.209 (0.230)	0.218 (0.238)	0.051 (0.136)	0.040 (0.143)
Controls		X		X
$R^2$	0.009	0.162	0.001	0.034
Number of observations	107	107	310	310
Dep. Var.: Index of friendships				
Race IAT	-0.029 (0.182)	-0.093 (0.206)	0.032 (0.118)	0.007 (0.118)
Controls		X		X
$R^2$	0.000	0.062	0.000	0.070
Number of observations	95	95	284	284

Notes: Standard errors are clustered at the room-level.

## D Impact on academic performance

In the main text of this paper, I have replicated and extended part of the research performed by Corno et al. [2022], that was connected to the following research-question: “Does engaging with individuals of a different race alter an individual’s stereotypes concerning that race?”. However, their paper studied a second main research-question: “What impact does inter-group contact have on academic performance?”

To answer the second research-question, I make use of several indicators of academic performance. Those include GPA, number of exams passed, whether the student is eligible to continue his or her study at the end of the academic year, and a general index of academic performance, constructed by Principal Component analysis. In order to use this data to answer the second sub-question, I again make use of linear regressions. It turns out there are significant effects of inter-group contact on academic achievement. This varied among different groups. For Black students who shared a room with non-Black students, their GPA scores improved by approximately 0.26 standard deviations. This improvement accounts for more than a fourth of the GPA gap between White and Black respondents. Furthermore, Black students in mixed rooms performed better in terms of passing a higher number of exams and being more likely to be eligible for the following academic year. Importantly, the positive effect on academic performance endured beyond the first year, even when most students were no longer residing in university residences.

In contrast, I did not find any significant impact on academic performance for White students in mixed rooms. The estimated effects for them were not significantly different from zero. Notably, the positive effect observed among Black students was not only caused by academic peer effects. In the used regressions, roommate’s entry score is used as a control for academic ability, and it typically showed no significant influence. Additionally, factors such as being in the same faculty or taking the same courses did not account for the positive performance effects observed among Black students in mixed rooms.

The following sub-sections will go into greater detail about the measures of performance, the used models, results and conclusions regarding the second sub-question.

### D.1 Measures of Performance

The first measure of academic performance is based on the Admission Point Score. This score is based on a student’s performance in high school. It is constructed by taking the average of the scores for English and five other courses in the final year off high school. Exam scores are on a scale from 0 to 100. For this research, I use a re-scaled version of this by dividing it by 1000. This re-scaled score is denoted as UCT entry score. This variable is used as an indication of academic ability at the start of a student’s time at UCT. Additionally, I use several measures of academic ability at the end of the first and second year. The first is the GPA. This score is standardized to have mean zero and standard deviation equal to one for all years. Furthermore, I use the total number of exams passed and failed, and the assessment of exam committees on whether a student is allowed to advance to the next year. The latter is a binary variable taking value one when a students is eligible to continue, and zero else. I refer to this variable as

“eligible to continue”. All measures mentioned in this sub-section were available from the UCT administry.

## D.2 Methodology

In order to estimate the effect of interracial contact on academic performance, I specify the following model:

$$Y_{ijk1} = \beta_1 MixedRoom_{ik0} + \beta_2 UCT_i + \beta_3 UCT_j + \gamma_1 X_{ik0} + \gamma_2 X_{jk0} + \gamma_3 \delta_k + \epsilon_{ijk1} \quad (10)$$

Here,  $Y_{ijk1}$  denotes the outcome variable of interest, which in this case are GPA, Number of exams passed, Eligible to continue and Index of academic performance.  $MixedRoom_{ik0}$  again denotes a binary variable taking value one when person  $i$  was paired with someone of another race.  $UCT_i$  and  $UCT_j$  are control variables controlling for own UCT entry score and roommates’ UCT entry score, respectively.  $X_{ik0}$  and  $X_{jk0}$  denote variables containing own and roommate controls, respectively. Controls include the following characteristics: household wealth, the student’s monthly consumption expenditure, and a dummy variable indicating whether the respondent is not from South Africa.  $\delta_k$  includes seven residence dummies. Finally,  $\epsilon_{ijk1}$  denotes the error term.

## D.3 Results

In Table 14, one can see the regression results for the four different outcome variables. For GPA, being in a mixed room has a significant positive effect for the Black sub-sample. The size of the effect (0.257) corresponds to approximately one third of the gap in GPA between White and Black students. For the White sub-sample and the full sample, the effect is not significantly different from zero. For the number of exams passed, I again find a significantly positive effect of being in a mixed room for Black students. On average, Black students in mixed rooms pass 0.645 exams more than Black students in same race rooms. For the White sub-sample, again no significant effect can be found. For the full sample, The effect is positive and significant, with a smaller coefficient than for the Black sub-sample. Next, I consider the binary variable “eligible to continue” taking value one when the student is allowed to advance to the next year of their study. Again, Black students in mixed rooms are more likely to advance to the next year than Black students in same race rooms, with the magnitude of the effect being 0.152. For the White sub-sample, the effect is again not significantly different from zero, The effect for the full sample is again positive and significant, which is driven by the positive effect for the Black sample. Finally, for the global index of performance, I again find a significant positive effect for being in a mixed room for Black students. The effect for the White sub-sample is again not significantly different from zero. Table 15 shows results for measures of academic performance in the second year as outcome variables. Note that those outcome variables are not cumulative, but are exclusively measured in the second year, while the variable Mixed Room still represents whether the student was in a mixed race room in the first year. As can be seen, the positive effect on GPA for the Black sub-sample seems to fade away. However, the effects on number of exams passed and eligible to continue remain positive and significant, and of similar size. Note that the effects are controlled for roommate ability and whether paired students were in the same faculty.

Table 14: Impact on academic performance in first year

	White			
	GPA	Num. of Exams Passed	Eligible to Continue	Perf. Index
Mixed Room	-0.028 (0.243) [0.922]	-0.168 (0.523) [0.922]	0.050 (0.066) [0.782]	0.010 (0.259)
UCT entry score	10.812*** (2.687)	10.553* (5.881)	-0.453 (0.634)	7.702*** (2.090)
Roommate UCT entry score	2.977 (1.871)	2.619 (5.210)	0.630 (0.625)	2.943 (2.135)
Mean of dep. var.	0.663	6.500	0.923	0.709
$R^2$	0.576	0.727	0.436	0.426
Number of observations	117	117	117	117
	Black			
Mixed Room	0.257** (0.125) [0.041]	0.645*** (0.245) [0.015]	0.152*** (0.040) [0.000]	0.443*** (0.141)
UCT entry score	5.505*** (1.392)	11.441*** (2.841)	0.738 (0.469)	6.158*** (1.618)
Roommate UCT entry score	1.111 (1.234)	0.844 (2.345)	0.444 (0.388)	1.316 (1.347)
Mean of dep. var.	-0.268	4.506	0.852	-0.281
$R^2$	0.386	0.715	0.400	0.447
Number of observations	332	332	332	332
	Full			
Mixed Room	0.147 (0.102) [0.147]	0.447** (0.204) [0.050]	0.105*** (0.031) [0.003]	0.289** (0.113)
UCT entry score	8.237*** (1.204)	12.750*** (2.346)	0.840** (0.387)	8.105*** (1.320)
Roommate UCT entry score	0.924 (0.957)	1.646 (1.862)	0.603** (0.295)	1.539 (1.021)
Mean of dep. var.	-0.044	4.977	0.871	-0.042
$R^2$	0.424	0.709	0.325	0.447
Number of observations	499	499	498	498
Controls	X	X	X	X
Roommate Controls	X	X	X	X
Academic program fixed effects	X	X	X	X

Notes: Standard errors in parentheses are clustered at the room-level. P-values in square brackets are adjusted for multiple hypothesis testing using the resampling method of Westfall and Young [1993].

Table 15: Impact on academic performance in year 2

	GPA	Number of exams passed	Eligible to continue	Index of academic performance
<b>Whites</b>				
Mixed Room	-0.082 (0.085)	0.428 (0.808)	-0.033 (0.070)	-0.073 (0.252)
UCT entry score	7.113*** (1.066)	-2.844 (9.152)	0.037 (1.055)	9.054*** (2.469)
Roommate UCT entry score	1.318* (0.768)	6.750 (7.272)	-0.421 (0.680)	2.475 (2.234)
<i>Mean of dependent variable</i>	0.683	6.099	0.958	0.570
<i>R</i> <sup>2</sup>	0.813	0.631	0.513	0.657
Number of observations	105	105	105	105
<b>Blacks</b>				
Mixed Room	0.052 (0.070)	1.012** (0.408)	0.150** (0.062)	0.451** (0.206)
UCT entry score	1.590** (0.706)	10.152** (5.071)	1.004 (0.668)	5.310** (2.135)
Roommate's UCT entry score	1.090** (0.548)	5.793* (3.331)	-0.300 (0.414)	2.148 (1.442)
<i>Mean of dependent variable</i>	0.254	4.785	0.879	-0.349
<i>R</i> <sup>2</sup>	0.444	0.596	0.412	0.476
Number of observations	208	208	207	207
<b>Full</b>				
Mixed Room	-0.016 (0.052)	0.805*** (0.299)	0.089** (0.039)	0.242* (0.132)
UCT entry score	3.467*** (0.608)	6.604* (3.909)	0.838* (0.482)	6.954*** (1.615)
Roommate UCT entry score	0.739* (0.408)	7.255** (2.908)	0.013 (0.296)	2.353** (1.095)
<i>Mean of dependent variable</i>	0.389	5.234	0.902	-0.055
<i>R</i> <sup>2</sup>	0.522	0.518	0.317	0.474
Number of observations	355	355	354	354
Controls	X	X	X	X
Roommate Controls	X	X	X	X
Academic program FE	X	X	X	X

Notes: Standard Errors in parentheses are clustered at the room level.

## **E Attitudinal and behavioural indices**

The survey questions were on three topics. The first topic is friendship patterns. Respondents were asked how often they socialized with individuals of other races, how many individuals of another race they preferred to have in a hypothetical seven-members study or leisure group, how many actual friends are Black/White and how many actual study partners are Black/White. The second topic is on interracial attitudes. Questions included: the frequency and comfort in discussing issues of race, agreement with abolishing affirmative action in college admissions, feelings of self-consciousness when dancing or dating with someone from another race. The last topic was on prosocial behaviour. Respondents were asked if they were a member of any community service or volunteer organizations and how much money they donated to charities in the last year, churches excluded.

For the regressions regarding explicit attitudes and behaviours, 15 measures were grouped into three indices and used as outcome variables. For the friendship index, the following measures were included: (i) the number of times the respondent socialized with someone of a different race in the last month (excluding the roommate); (ii) the last time the respondent socialized with people of a different race; (iii) the fraction of actual friends and study mates of a different race (excluding the roommate); (iv) the number of desired friends of a different race in a hypothetical leisure group and in a hypothetical study group (excluding the roommate). For the index of explicit attitudes, the following were included: (i) the frequency with which the respondent talked about discrimination and racial bias with friends in the last month;<sup>45</sup> (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.” And finally, for the index of prosocial behaviour, I included: (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.

## **F Residence choices**

Table 16 shows the results for regressions regarding residence choices at the end of the first year.

Table 16: Residential Choice at the end of the first year

	Whites	Blacks	Full sample
Dep. variable: still in residence in year 2 1 if in mixed Room at baseline	0.014 (0.110)	0.061 (0.055)	0.073 (0.048)
Mean of dep. var.	0.744	0.786	0.763
$R^2$	0.134	0.084	0.118
Number of observations	117	332	499
Dep. variable: Still in residence and in mixed room in year 2 Mixed Room	0.084 (0.058)	0.009 (0.028)	0.039 (0.029)
Mean of dep. var.	0.013	0.023	0.020
$R^2$	0.158	0.056	0.052
Number of observations	117	332	499
Dep. variable: Still in residence and same roommate in year 2 Mixed Room	0.038 (0.037)	0.023 (0.026)	0.019 (0.027)
Mean of dep. var.	0.026	0.027	0.026
$R^2$	0.103	0.053	0.037
Number of observations	117	332	499

Notes: Standard errors in parentheses are clustered at the room-level.

## G Programming code

### G.1 Replication

In order to obtain the replicated results of Corno et al. [2022], I largely made use of the replication package provided. In the submitted files, I have changed the following things compared to the original code:

1. I have added the following piece of code in the 0\_main file to be able to utilize the right ado-file:  

```
sysdir
adopath + "C:\Users\abeke\Downloads\174501-V1 (2)\ForJournals_a\programs\ado"
```
2. In the codes for the adjusted P-values for Table 4 and Table A9 of the original paper, I have changed the variable `regprogram_fwer` to `regprogram` as the variable `regprogram_fwer` is non-existent in the dataset and not generated by the code. This is just a small error by the authors.
3. In the do-files for the adjusted P-values for all tables, I have set all variables *full*, *White* and *Black* equal to one. In the original code, some were set equal to zero, which prevented their respective parts of code to be executed by means of an if-statement.

### G.2 Extensions

In order to obtain the results for my extensions using roommate IAT scores, I have first created an R-script file called `roIAT.R` to create a variable representing roommate Race IAT score, as this was not present in the original dataset. I loaded the file `uctdata_balanced.dta` in Rstudio and I have obtained the roommate IAT scores by matching students based on their room number



at baseline, using a double for-loop. I also obtained a dummy variable taking value one when the student was more prejudiced towards Blacks than the corresponding roommate. Additionally, I re-scaled the Roommate UCT entry scores to have mean zero and standard deviation equal to one. I then wrote the obtained dataset as a .dta file, to be able to use in Stata. I used the file IATextensions\_final.do for the extensions on own baseline IAT and roommate IAT. I used the file roUCTextensions\_final.do to obtain the estimates for the models related to heterogeneity by roommate UCT score.

All codes were run using R version 4.3.0 and Stata 17