Different behaviors in College Applications of High-achieving students

Alice Battiston, Luca Perdoni

July 16, 2015

1 Introduction

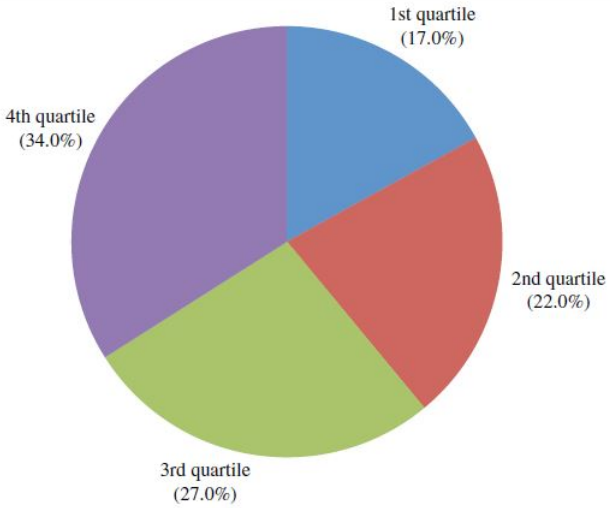
We want to build a model in NetLogo that simulates the application choices of high-achieving students. In order to do so, we base our understanding of the issue mainly on *The Missing "One-Offs": The Hidden Supply of High-Achieving, Low Income Students* by Caroline M. Hoxby (Stanford University) and Christopher Avery (Harvard Kennedy School), which deeply analyses the subject in the United States. All pictures in this section are drawn from their paper.

First of all we must be very clear about who are the *agents* we are interested in, since they are a very tiny portion of high school students population. All our agents will be very well prepared students who are likely to be admitted to selective institutions if they apply; according to the paper this group represents approximately 4% of U.S. students. Of course, such a small fraction of students will have peculiar features, different from the whole population. From figure 1 we can see as "rich" students , defined as the one with a family income in the 4th quartile of the U.S. annul family income i.e. above 120,776 $, are over represented in the high-achieving students category while low-income students, those who are in the first quartile i.e. below 41,472 $ are under-represented. In our model we will consider only agent in the 1st and in the 4th quartile, being coherent with the relative weights in terms of population.

Another important feature is where these students live, which can be easily seen in Figure 2. Critical masses of high-achieving students can be found in Massachusetts, Connecticut, Rhode Island New York, New Jersey, eastern Pennsylvania, while another group is in southern Florida, the other concentration is in coastal California from the Bay Area to San Diego. Other

Figure 1:

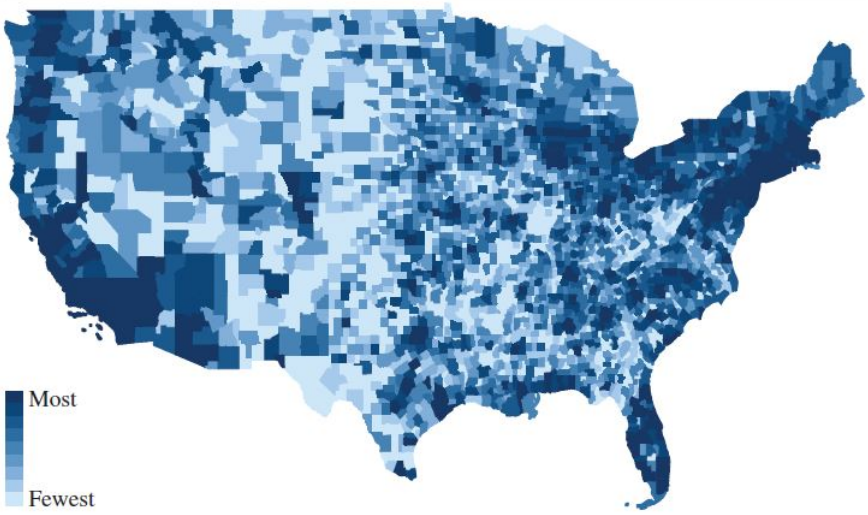
High-Achieving Students, by Family Income Quartile[a]



Source: 2008 American Community Survey and authors' calculations using the combined data set described in the text.

Figure 2:

Numbers of High-Achieving Students, by County[a]

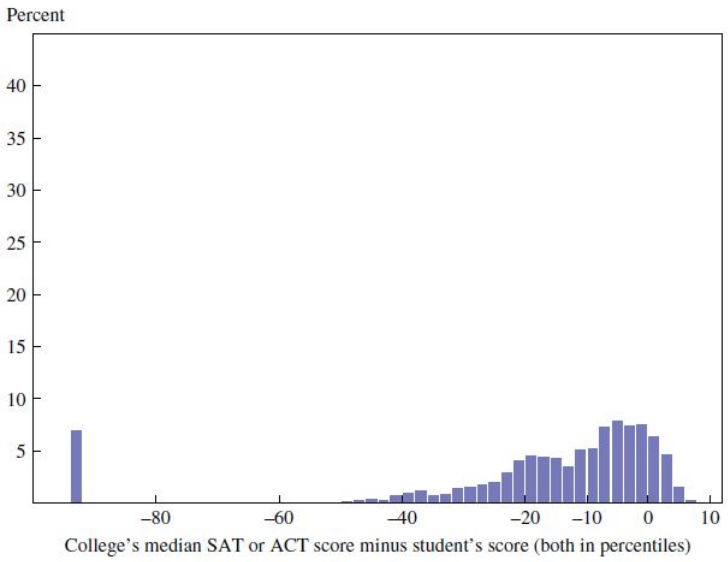


Source: Authors' calculations using the combined data set described in the text.

a. Counties are ranked by the absolute number of high achievers (as defined in table 2) living in the county in 2008 and then grouped into deciles; counties are then shaded according to their decile.

Figure 3:

Distribution of High-Achieving, High-Income Students' College Applications, by Student-College Match[a]



Source: Authors' calculations using the combined data set described in the text.

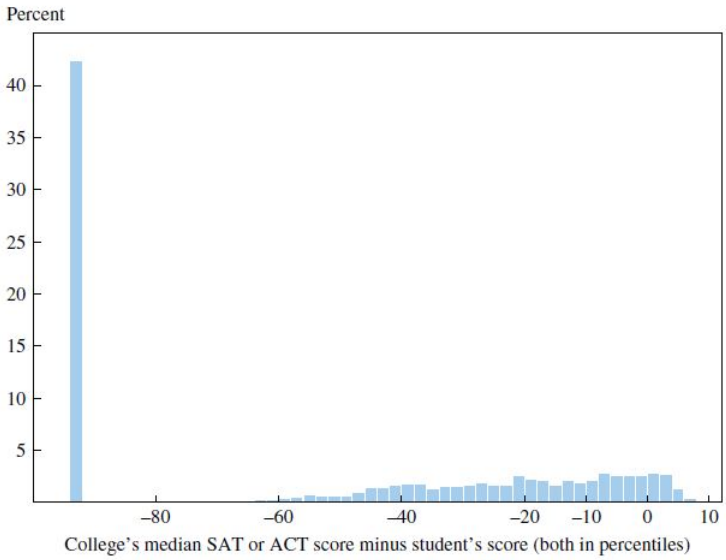
critical masses are more scattered as in Chicago, Houston, Dallas-Fort Worth, Atlanta.

We would expect a homogeneous behavior in college applications from such a small group of people, but it turns out that this is not the case. This can be clear if we look at Figure 3 and 4 which illustrate the distribution of college applications in terms of difference between college's median standardised test score and the student's own. Since we are dealing only with excellent students, colleges that correspond to values between -5 and +10 are excellent institutions while the ones below -10 represent safe choices. We must also consider applications to non-selective institutions which do not require standardised tests and where everybody can enrol. If we look at Figure 3 we can see the behaviour of high-income high-achieving students which is totally consistent with expert counsellor's advices. They apply to very prestigious institutions plus some safe applications which are due to the very low admission rate of those colleges. The percentage of non-selective applications, the bar on the left, is very low and in this case refers to music conservatories, design school or drama and performing arts school. We will refer to this college application method as *achievement-typical*.

A completely different situation can be observed in Figure 4 where the applications of low-income students are plotted. The most striking feature is

Figure 4:

Distribution of High-Achieving, Low-Income Students' College Applications, by Student-College Match[a]



Source: Authors' calculations using the combined data set described in the text.

the huge percentage of non-selective applications which goes together with a behaviour similar to the achievement typical one previously described. Hoxby and Avery found that 8% of low-income students follow an achievement typical behaviour while 53% do not apply to school within 15 percentiles of their own even though they apply to at least one non selective college. We will refer to this college application method as *income-typical*. Note that 39% of low income students follow neither of the two typical behavior and apply with a logic which is difficult to identify.

In our simulation we will try to model a world where high-income students apply according to achievement typical method while low- income students split into two groups. The choice of which scheme to follow when applying to colleges by low-income students is the crucial part of our simulation. It follows that the main question is why a low income student should behave in one way with respect to the other. In order to answer this question is useful to look at differences in income, race, community types and High school features. Figure 5 shows that income-typical students are, in average, richer that achievement-typical and their choice can hardly be caused by a socio-economic disadvantage due to their race. Figure 6 instead gives a strong hint on what could be one of the main reason driving such a different behaviour. A great part (65%) of achievement typical students live in metropolitan areas

Figure 5:

Socioeconomic Characteristics of High-Achieving Students[a]

Characteristic	*High-income students*	*Low-income students*	
		Achievement-typical	*Income-typical*
Annual family income (dollars)[b]	157,569	30,475	32,418
Parents' education (years)[c]	18.7	16.0	16.7
Race or ethnicity[d] (percent of total)			
White	74.8	45.1	79.5
Black	2.1	5.2	2.9
Hispanic	5.6	12.6	6.0
Asian	20.5	31.8	7.3

Source: Authors' calculations using the combined data set described in the text.

Figure 6:

Types of Communities Where High-Achieving Students Reside

Percent

Community type	*High-income students*	*Low-income students*	
		Achievement-typical	*Income-typical*
Main city, urban area with population > 250,000	17	26	8
Main city, urban area with population 100,000–250,000	14	21	13
Main city, urban area with population < 100,000	48	18	9
Suburb, urban area with population > 250,000	8	9	9
Suburb, urban area with population 100,000–250,000	0	2	2
Suburb, urban area with population < 100,000	0	4	12
Town, near an urban area	0	5	12
Town, far from an urban area	5	7	13
Rural, near an urban area	6	4	10
Rural, far from an urban area	0	5	10

Source: Authors' calculations using the combined data set described in the text.

while only 21% live in non-urban areas, which is exactly the opposite picture with respect to income-typical who mainly live in little towns and rural areas. The main idea of the paper is that income-typical behaviour is due to a lack of information on very selective colleges. This hypothesis seems stronger if we look at Figure 7 where the main difference between the two groups is relevant in terms of previous student who attended a top college and also the physical miles needed to draw a circle gathering 20 or 50 high achievers. We will be coherent with these results in our simulation.

Figure 7:

College-Related Characteristics of High Schools Attended by High-Achieving Students[a]

		Low-income students	
Characteristic	*High-income students*	*Achievement-typical*	*Income-typical*
Percent of teachers who graduated from a peer college[b]	8.9	2.9	1.1
Percent of teachers who graduated from a safety college[c]	14.4	7.5	5.0
Number in a typical previous cohort who applied to top 10 U.S. colleges[d]	12.9	7.6	1.6
Number in a typical previous cohort who were admitted to a top 10 U.S. college[d]	12.3	7.4	1.5
Number in a typical previous cohort who enrolled at a top 10 U.S. college[d]	12.3	7.4	1.5
Percent of cohort who are high achievers	17.1	11.2	3.8
Radius to gather 20 high achievers (miles)	2.6	7.7	19.3
Radius to gather 50 high achievers (miles)	4.1	12.2	37.3

Source: Authors' calculations using the combined data set described in the text.

2 The Model

The purpose of the model is to simulate the behavior of the high-achieving students. We offer a basic model and an extension of it.

The basic model The basic model can be exploited to detect how the behavior of each applicant is influenced by social interaction. The fundamental assumptions of the basic model are:

- **Assumption 1- The world**
  The artificial world we set up is composed of 4 states. Each of these state has its own geographical feature: they differ in terms of the number of cities and extension of rural area.

- **Assumption 2- The universities**
  There exist 6 universities, which differ according to their ranking position. One university is a selective one, 4 universities are state-flagship institutions (medium ranking position) and the last one is a non-selective one.

- **Assumption 3- Behavior determination**
  All the high-income students behave as achievement-typical. Low-income applicants are initially all income-typical, but once they meet an AT students, they can be influenced and change their behavior. The higher the number of achievement-typical students they meet, the higher the probability of changing their behavior.

- **Assumption 4- Different degree of permeability to information**
  Applicants are not all equal: the degree of permeability to new informations differs among students.

- **Assumption 5- Relation income/hometown**
  High income students are born in a big town with 80% probability, suburban area with 10% probability or rural area with 10%. Conversely, low income students are born in a big town with 50% probability, suburban area with 20% probability and in a rural area with 30%. These distributions are inferred from the paper we refer to.

Extension: interaction with students admitted to a top university The extended model aims to describe the effect of students from a top university on applicants who are born close to their hometown. The simulation is

composed of two periods. A first period is run merely to create a generation of students who gain admission to a top university. The second period works exactly as the first one, except for one feature: now applicants can meet both students of a top university and other applicants. The main assumption is:

- **Assumption 6- Top university's students influence**
  All applicants display an higher permeability to information from a top-college-admitted student than from other applicants.

3 The code

In order to run the basic model, the user has to press only $SETUP$ and GO. In the basic model there are only applicants as agents and we can observe the interaction of students who are called to apply at the same time. If the user is interested in running the extended model, he has to proceed in the following way:

1. The basic model has to be run. Indeed, the first period of the extension corresponds to the basic model and the results of it are used to identify the students who gained admission to the top university.

2. Use $SETUP{-}SECOND{-}PERIOD$ and $GO{-}SECOND{-}PERIOD$ to run the second period.

In what follow we will describe the main commands used in the basic model (see *First Period*) and in the extended model (see *First Period + Second Period*).

3.1 First Period

Setting up the world The world is divided into 4 regions (N-E, N-W, S-E, S-W). These territories share some common features as:

- Each region has at least one *main town* (red patches) with its *suburban area* (grey patches).

- Each region has at least one university: the *state-flagship university.*

Setting up the world has required the use of a number of commands, among those the main one is *breed.* The command *Breed* allows us to differentiate between students and universities, specifying their own attributes.

```
universities-own [ranking-position
  applications
  State-Uni]

globals [second-period
  ]

turtles-own [period]

links-own [chooser]

students-own [initial-coor
  income
  hometown
  InitialBehavior
  FinalBehavior
  threshold
  state
  d1
  d2
  contact
  e1
  e2]
```

Students display a high number of attributes. They own:

1. *Period* This attribute allows to distinguish between students born in the first period of the simulation and students born in the second period.

2. *Income* Income is attributed randomly between 0 and 100.

3. *Hometown* Hometown is used to detect in which area (*Big*, *Suburban* or *Rural*) students are born. First students are asked to set their coordinates randomly, then they are distributed according to *assumption 5* through the command *move-to*.

```
ask students [let i random-float 1 if income > %ofLowIncomeBetweenHighAchivingStudents
  and period = 2 [if i <= 0.1 [move-to one-of patches with [pcolor = 55]]
    if (i > 0.1 and i <= 0.2) [move-to one-of patches with [pcolor = 8]]
    if (i > 0.2) [move-to one-of patches with [pcolor = 15]]
    set initial-coor list (xcor) (ycor)]]
 ask students [let i random-float 1 if income <= %ofLowIncomeBetweenHighAchivingStudents
  and period = 2 [if i <= 0.3 [move-to one-of patches with [pcolor = 55]]
    if (i > 0.3 and i <= 0.5) [move-to one-of patches with [pcolor = 8]]
    if (i > 0.5) [move-to one-of patches with [pcolor = 15]]
    set initial-coor list (xcor) (ycor)]]
```

4. *InitialBehavior* It allows to detect the behavior before students start talking and exchanging information. Its setting respects *assumption 3*.

5. *FinalBehavior* It describes the final behavior of students. At the beginning $FinalBehavior$ is set equal to $InitialBehavior$.

6. *Threshold* It allows to differentiate the degree of students' permeability to contact with the other agents. Each student is asked to set his threshold randomly between 0 and 1.

7. *Initial-coor* This attribute is merely used to recall their initial position. Each simulation lasts two periods in order to verify the effect of students previously admitted to top universities to applicants' behavior. Indeed, those first period's students who gain admission to the top university go back to their initial coordinates at the beginning of the second period sharing information with the new applicants.

8. *d1 & d2* These attributes are variable taking values 0 or 1. They will be used in the second period of each simulation in order to detect whether the new students are born close to students who have been already admitted to a top university. Two sliders allow to set the desired distances.

9. *e1 & e2* As the previous attributes, e1 and e2 take values 0 or 1. They can be used in any period in order to detect whether the new students are born close to other agents or not. Reference distances are set with sliders $distance1$ and $distance2$ (the same for d1 and d2).

```
ask students [if period = 2 and count(turtles with [ period = 1 ] in-radius distance1) > 1
   [set e1 1 ]]
ask students [if period = 2 and count(turtles with [ period = 1 ] in-radius distance2) > 1
  and count(students with [ period = 1 ] in-radius distance1) <= 1
  [set e2 1 ]]
```

10. *contact* It signals how many of the potential influencer agents each student meets.

Conversely, universities own only three attributes:

1. *Ranking-position* Universities are classified according to their selectivity. One university is a top one, one is non-selective and the remaining four are flag universities.

2. *State* This attribute detects which state each university belongs to.

3. *Applications* It detects the number of application each university receives.

Finally, the only attribute of links is a chooser. Links are created to represent students' applications. In the default simulation they are hidden, but as the switch is turned on they gain a new visible shape.
Some of the variables have to be chosen by the user. In particular, the user is called to set how long students have to share information ($\#ofTicks$) before applying to universities, how many students there are ($\# - of - students$) and which percentage of students is low income (*%ofLowIncomeBetween-HighAchivingStudents*) between the high achieving students (the only one we shown in the simulation). Notice that even though $distance1$ and $distance2$ can already be set, they are not meaningful in the first period.

Making the simulation run The structure of the go button is simple. At any tick students move, they could share information and, thus, modify their behavior. As the time is over (the limit is set by $\#ofTicks$) students apply to the universities according to their $FinalBehavior$.

How do students share informations?

```
to share-info
  if ticks > #ofTicks [stop]
  ask students [if InitialBehavior = "IT"
      and count(students with [InitialBehavior = "AT"] in-radius 0.5) > 1 [set contact contact + 1]]
  ask students
    [if InitialBehavior = "IT"
      and count(students with [InitialBehavior = "AT"] in-radius 0.5) > 1
      and random-float 1 > threshold
        [set FinalBehavior "AT"]]
  ask students
    [if InitialBehavior = "IT"
      and count(students with [InitialBehavior = "AT"] in-radius 0.5) > 1
      and threshold >= 0.1
        [set threshold threshold - 0.05]]
end
```

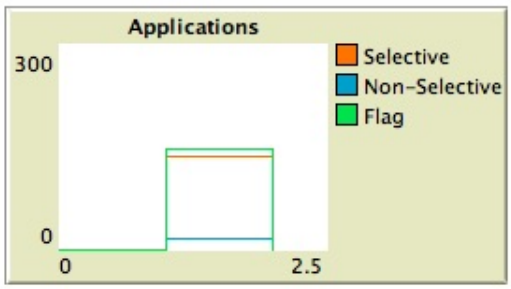
As agents move, it could be the case that students, whose $InitialBehavior$ is income-typical, get in touch with some achievement-typical students. Getting in touch means that at least one of the students whose FinalBehavior is achievement-typical lies in a circle of radius 0.5 around the income-typical student. Once two students with the above characteristics meet, the behavior of the income-typical changes with a probability of $(1 - threshold)\%$. The change in the behavior is monitored updating $FinalBehavior$ to AT. Moreover, since one of the assumptions of the model we built up is that meeting more students implies having a higher probability of being influenced. Thus, every time a income-typical student meets a achievement-typical one, the

value of his threshold falls down of 0.05 (as long as it is greater or equal than 0.1). The proportion of income-typical students who change their behavior is shown in a monitor.

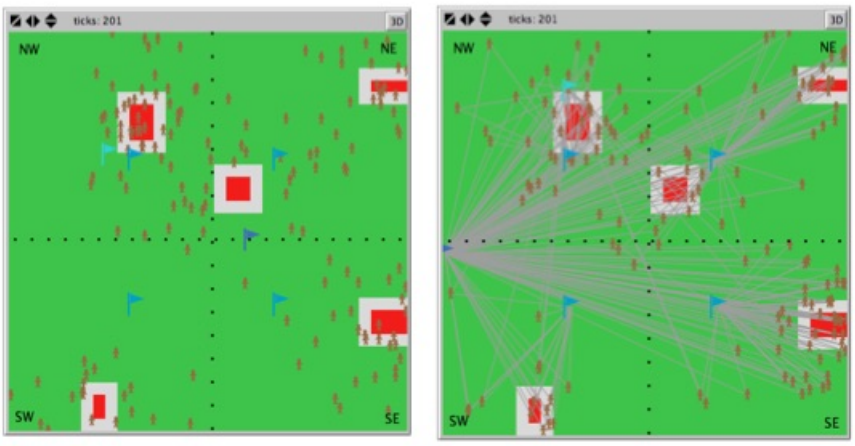
How do students apply to the universities?

```
to application
  ask students [ if ticks = #ofTicks and FinalBehavior = "AT"
    [ create-links-with universities with [ranking-position = "selective"]]
      if ticks = #ofTicks and FinalBehavior = "IT"
      [create-links-with universities with [ranking-position = "non-selective"]]
      if ticks = #ofTicks and state = "NE"
      [create-links-with universities with [State-uni = "NE"]]
      if ticks = #ofTicks and state = "SW"
      [create-links-with universities with [State-uni = "SW"]]
      if ticks = #ofTicks and state = "SE"
      [create-links-with universities with [State-uni = "SE"]]
      if ticks = #ofTicks and state = "NW"
      [create-links-with universities with [State-uni = "NW"]]]
  ask universities [set applications count my-links]
  ask links [ifelse Shape-links [if chooser = 1 [set shape "default"]]
    [set shape "trasparenti" ]]
end
```

Once $ticks = \#ofTicks$, it's time to apply! Applications depends only on the $FinalBehavior$ of each student. Income-typical students apply to the non-selective university and to their state's flag university, while the achievement-typical ones send their applications to the top university and to the flag university of their state. Applying to a university in our model means to create a link with it. Indeed, all students whose $FinalBehavior$ is AT are asked to create a link with the top university and with the flag university, whose state is equal to the student's state; conversely, income-typical students create a link with the non-selective university as well as with their flag university. Thus, universities can check the number of applications they have received by counting their links. Applications to each class of universities is shown by mean of a graph.



An important remark is that since everyone applies at its flag university, the green line always report the number of students (applicants). Links can be invisible or visible, the visible-mode has to be switched on before the simulation starts.



3.2 Second Period

Setting up a new generation As the first period ends, all the students who didn't gain admission to the top university die. Indeed, students who have been admitted to the top university are assumed to be more *powerful* than the standard agent when meet new applicants. How does the admission process work?

```
ask students [if period = 1 and FinalBehavior = "IT" [die]
  if period = 1 and FinalBehavior = "AT" and random-float 1 <= 0.8 [die]]
ask students [if period = 1 [let p first [initial-coor] of self
  let new-coor reverse [initial-coor] of self
  let q first new-coor
  move-to patch p q]]
```

All the students we monitor are assumed to be high achieving students, hence, the code select randomly the admitted students between those who applied. Indeed, each student who was born in period 1 and whose $FinalBehavior$ is AT has a 20% probability of being admitted. Notice that what the code is required to do is to kill those students who have been rejected by the top university. Then the lucky admitted students are asked to go back to their initial position. How does it happens? Recall that through the attribute $Initialcoor$ each student memorized its initial position. The command $first$ allows to extract the first element of the vector $Initialcoor$, i.e. the x-coordinate (p). Then reversing the vector and applying again the $first$ command we can

extract the y-coordinate (q) too. Finally we ask each survived student to move to the patch with coordinates p and q.

Finally a new generation is born with the same characteristics of the first one , but the two attributes $d1$ and $d2$.

```
ask students [if period = 2 and count(turtles with [ period = 1 ] in-radius distance1) >= 1
   [set d1 1 ]]
ask students [if period = 2 and count(turtles with [ period = 1 ] in-radius distance2) >= 1
  and count(students with [ period = 1 ] in-radius distance1) < 1
  [set d2 1 ]]
```

Note that $d1$ and $d2$ take values 1 or 0. The attribute $d1$ is equal to 1 if at least one student of the previous period lie in the circle of radius $distance1$ around the student, otherwise it is 0. Conversely, the attribute $d2$ takes value 1 if no students of the previous period lies in the circle of radius $distance1$ around the student but at least one student of the previous period can be found in the circle of radius $distance2$.
All the other attributes are set as in the first $SETUP$, hence, the code is the same, but the specification *if period = 2* before any command.

Let's go (again)! Here, again, the code is the same as for the first *go*. The main difference is that now students can meet some students of the previous period. Hence, the new command is:

```
to share-info-bis
  if ticks > #ofTicks [stop]
  ask students [if InitialBehavior = "IT" and period = 2
    and (count(students with [InitialBehavior = "AT" and period = 2] in-radius 0.5) > 1 or
      count(turtles with [period = 1] in-radius 0.5) > 1 )
    [set contact contact + 1]]
  ask students [if InitialBehavior = "IT" and period = 2
    and count(students with [InitialBehavior = "AT" and period = 2] in-radius 0.5) > 1
    and random-float 1 > threshold [set FinalBehavior "AT"]]
  ask students [if InitialBehavior = "IT" and period = 2
    and count(students with [InitialBehavior = "AT" and period = 2] in-radius 0.5) > 1
    and threshold >= 0.1 [set threshold threshold - 0.05]]
  ask students [if InitialBehavior = "IT" and period = 2
    and count(turtles with [period = 1] in-radius 0.5) > 1
    and random-float 1 + 0.2 > threshold [set FinalBehavior "AT"]]
  ask students [if InitialBehavior = "IT" and period = 2
    and count(turtles with [period = 1] in-radius 0.5) > 1
    and threshold >= 0.1 [set threshold threshold - 0.05]]
end
```

The first 6 lines of the code are the same as in *to share-info*, but clearly they only refer to the new students (indeed, the check *If period = 2* is there). The difference lies in the second part of the code. Now income-typical students

can also meet with admitted students. If it is the case, the probability of changing behavior is higher than when meeting an achievement typical applicant. Indeed, now the probability is given by $1 - (threshold - 0.2)\%$.

4 Experiments

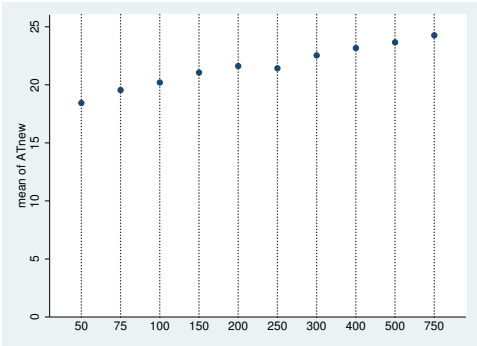
In this section we analyse the results we obtained simulating many different environments with the tool *BehaviorSpace* endowed in NetLogo. We are going to present many different experiments were we looked for the effect of several different variables. It must be noticed that, if not otherwise stated, we always start from a standard simulation with the following setting of parameters:

- $\#ofTicks = 200$
- $\# - of - students = 100$
- $\%ofLowIncomeBetweenHighAchivingStudents = 40$

This choice of parameters guarantees a *neutral* setting for IT students; In fact , taking the average, in this simulation an IT applicant changes behavior half of the times. Considering only the percentage of low-income students, it might seem wrong that there are more high-income than low-income applicants, but this is an empirical evidence we draw from the Hoxby,Avery's paper and it is typical of high-achieving students.

4.1 Does information period length matter?

In this simple experiment we measured the number of applicants who started with a behavior IT and ended with an AT one with different values of time length (as number of ticks). We used as tick levels the following numbers: $50, 75, 100, 150, 200, 250, 300, 400, 500, 750$. For each number the simulation was repeated 200 times. In the following graph the mean of the number of IT students who changed their minds is plotted for every value previously listed.

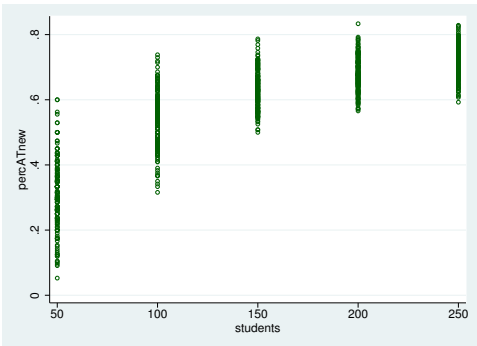


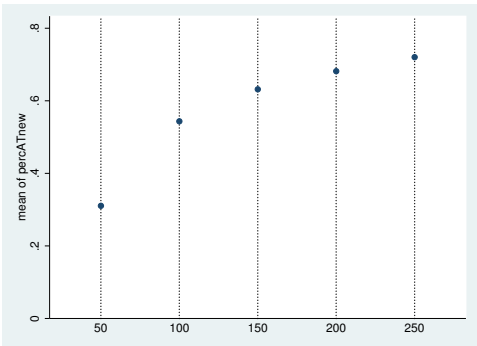
ticks	Summary of ATnew Mean	Std. Dev.
50	18.435233	3.6253164
75	19.549223	3.8969689
100	20.202073	3.6480011
150	21.055	3.3777241
200	21.615	3.790596
250	21.425	3.7449296
300	22.535	3.6627659
400	23.17	3.7871908
500	23.66	3.5563518
750	24.265	3.4048111
Total	21.614452	4.0464209

As it can be clearly seen there is a positive effect of the time length growth on the students who benefit from the exchange of informations. This result is coherent with our model (*assumption 3*) since the more time you have, the more people you meet, the higher the chance you change your behavior.

4.2 Does the number of students matter?

In this experiment we measured the ratio between the number of students who started as IT and finished as AT and the total initial number of IT. We collected this percentage varying the total number of students involved in the simulation, in order to understand whether this parameter affects the final outcome or not.

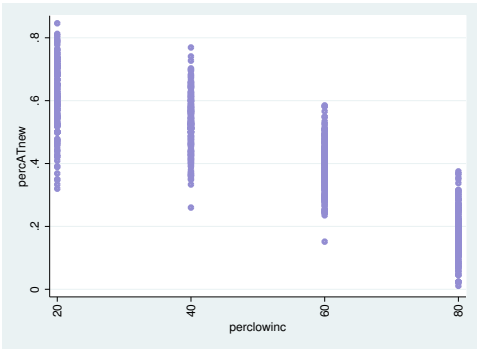


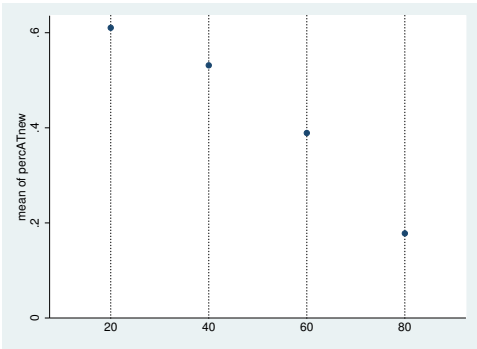


In the left graph we can see the dispersion of the percentage of students who changed behavior , while in the right picture we have the mean for every level of student's number. It is clear from the images that a more crowded world leads to much more students who becomes AT students. This is because of our assumption 3 that the choices of applicants are deeply affected by the people they met, and of course an increase in our variables implies a greater chance of speaking with other students.

4.3 Does the percentage of low income students matter?

In this experiment we are going to analyse the effect of a change in the percentage of low-income population over the total number of IT applicants who become AT. Accordingly to our assumptions, we expect that an increase in the size of the low-income group will lead to a fall in the fraction of IT students who are influenced by the interaction with other agents. Our predictions have been confirmed by the data collected with NetLogo, which are shown below





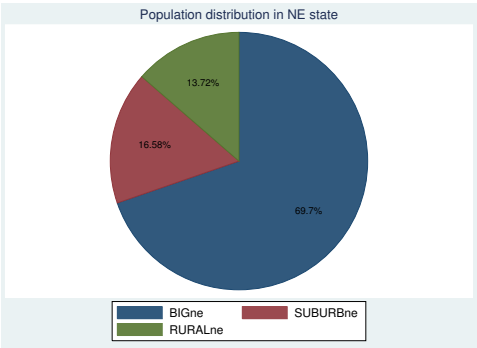
When we simulate an environment with only 20% of low-income students, and therefore a 0.8 share of high-income/AT students, we obtain that, in average, 61% of IT students change their behavior, while only 18% of them do it when the income distribution of the population presents 80% of low-income students.

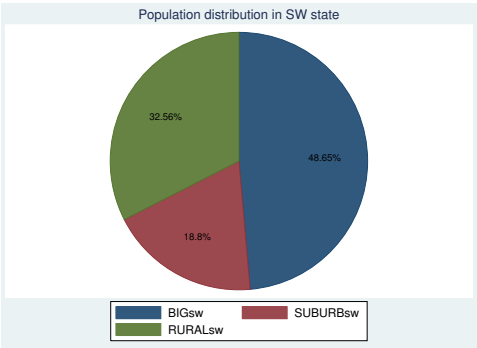
Can a income policy be effective? Observing this experiment we might wonder if a policy that increases IT students' income, therefore decreasing the percentage of low-income population, could affect the fraction of IT students who change their minds. We should not be very hopeful in expecting such results from such a policy. Why? As also stated by Hoxby and Avery, it is not purely the financial factor that leads to a IT choice. Agent's income determine where they grow up, who they meet and then how they see their future. Since we are considering only high-achieving students in a US-like setting we know that top universities would provide very generous grants to a low-income students once admitted: it is not that they can not, many do not want to apply to a selective institution. Our proposed policy could be effective only in the sense of changing the socio-economic environment, leading to a higher number of already AT students on the world. Therefore

the effects would not appear immediately, it would take at least a generation, and so many other factors could affect a change in the income distribution of families. We can conclude this policy would be ineffective.

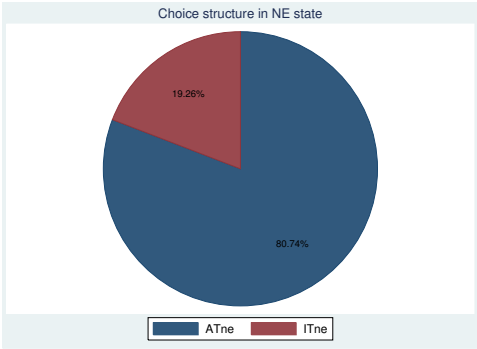
4.4 How urbanization can affect behavior

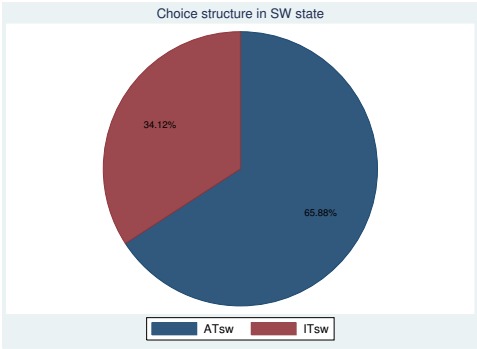
Before considering which are the feature of IT and AT applicants it might be useful to describe the world we created in NetLogo. In particular, due to our assumption, we expect that a different geography in a state may lead to a different choice structure in the state population. In order to understand if this is verified we performed 500 simulations in a standard setting. In the following we are going to compare NE and SW, since the first is the most urbanized state while the latter has the greatest percentage of rural area.





It is clear from the pie charts above that the greatest part of NE people live in a city (70%) while only 13% live in a rural area against the much bigger part (33%) of countryside inhabitant in SW. As already mentioned several times, a high-income agent has a much higher probability to live in a city with respect to a low-income one, in addiction to the fact that a rural student will have less social interactions with AT students. Given this assumptions, we expect a higher share of AT students in NE and a lower one in SW. Our thoughts seems confirmed from the results of the experiment, even if the difference is not as big as we might expect.





In fact, in NE the population presents 81% of AT while in SW the share decreases to 66%.

4.5 A picture of selective college applicants

It is interesting to understand what is the background of AT students, those who apply to a top college. In order to discover this, we collected data from 500 repeated simulations, with the standard setting mentioned before. Our results contain informations about the proportion of AT students which are high-income and the ones who are low-income, as well as the state where they come from and if they grew up in a city, in a suburb or in a rural area. Data are summarized in the table below, considering average values.

Variable	Obs	Mean	Std. Dev.
rich	500	.6947459	.0375839
poor	500	.3052541	.0375839

Variable	Obs	Mean	Std. Dev.
big	500	.7792929	.0369067
suburb	500	.1368189	.0294867
rural	500	.0838881	.0237903

Variable	Obs	Mean	Std. Dev.
NE	500	.3304313	.045587
SE	500	.2714381	.0398854
SW	500	.1127522	.0355471
NW	500	.2842019	.0438084

Our AT applicants have a strong characterization in terms of income: 70% of them is a high-income indeed. Recall that in our model an high-income agent directly starts with an AT behaviour and it never changes its mind towards

a IT choice. If we focus on the *hometown* of AT students we can see how the majority of them (77%) comes from a urban area, while only 13 out of 100 grew up in a suburb and only 8% of them came from the countryside. Again, remember that we assumed a precise matching between *income* and *hometown*: a high-income agent (i.e. an AT one) has a higher chance of living in a city rather than in a rural area. So we are not surprised that rural people are under-represented in AT applicants; this result agrees with similar results in Hoxby and Avery's paper.

It can be interesting to compare the share of AT students from the NE state with the share from SW. Why are they so different? It is not that strange if we take a look at the NetLogo world. The number of AT students, which is strongly influenced by income of the applicants, is a direct consequence of the geography of a state. The vast rural area plus the very little capitol city of the SW state implies a scarce presence of high-income people with respect to other states. States with a similar degree of urbanization, as NW and SE, reports similar shares in AT applicants indeed.

4.6 A picture of non-selective college applicants

As just analysed for AT students, we performed the same experiment for IT students. It has been used exactly the same procedure as before. Tables of the results are shown below. Note that we did not entered a table for wealth composition since all IT students, because of our model assumptions, are low-income.

	Mean	Std. Err.	[95% Conf.	Interval]
big	.1927354	.0046432	.1836129	.201858
suburban	.2050528	.0044948	.1962216	.2138839
rural	.6022118	.0056523	.5911067	.613317

	Mean	Std. Err.	[95% Conf.	Interval]
NE	.3135911	.0053837	.3030135	.3241687
NW	.2367955	.00499	.2269914	.2465996
SE	.2055567	.0044433	.1968268	.2142866
SW	.2343624	.0047738	.2249831	.2437416

The situation described in these tables is quite opposite to the one we had before. The majority of IT students comes from a rural area, while the remaining part is split equally between city and suburb. This comes from the fact that an agent starting in a green patch has far less chances to meet other applicants since they are more scattered. The table for the state provenance

might seem a bit puzzling but it is not. First of all note that the SW share, the state which has the greatest part of countryside, is much higher in IT students than in AT students, following our standard assumption. We could expect a lower share for NE, since it has less rural areas with respect to the other states, but we must consider the fact that city attracts much more people, independently from their wealth because of our set-up structure. The NE state then is the most populated state in our world: it is not strange that a third of IT students come from there.

4.7 Proximity as determinant of choices

Our assumptions suggest that a key factor in order to determine a behavioral change is who an agent meets. Since NetLogo allows for a spacial dimension, we thought that it would be interesting to measure the effect of proximity on choices of IT population. For this purpose we created variables *e1, e2, d1, d2* which are described in section 3. We ran an experiment in a standard setting with 500 repetitions and we measured the share of IT students who, when they were born, were very close to another agent (*e1*= 1), together with the share of IT students with *e2*=1 and the remaining part representing IT applicants such that *e1*=*e2*=0. We measured the same data for the group of students who started with an IT behavior and switched to a AT one. Results are shown below.

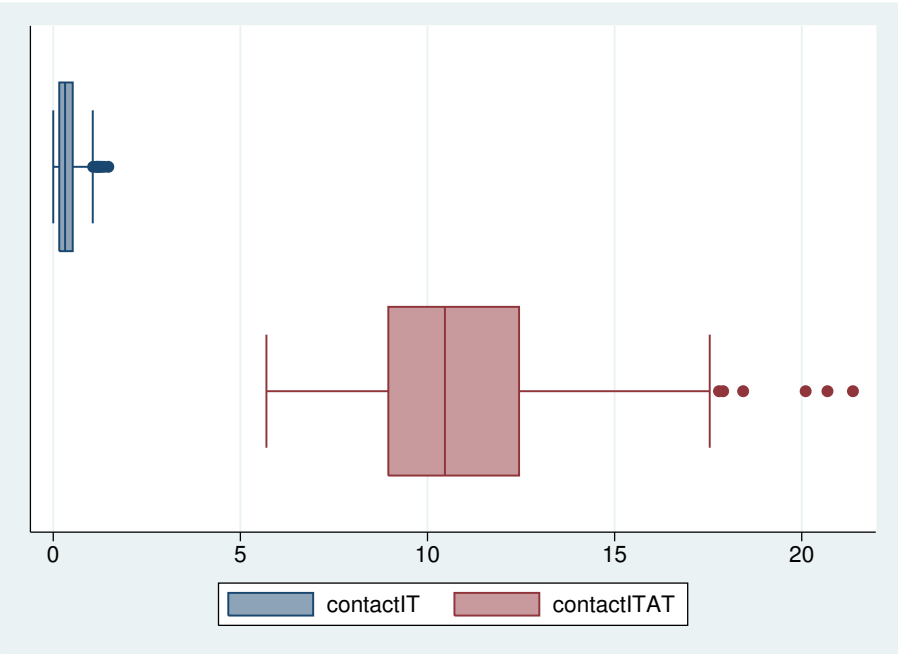
	Mean	Std. Err.	[95% Conf.	Interval]
ITe1	.6246895	.0052811	.6143136	.6350654
ITe2	.3442888	.005333	.3338109	.3547668
ITe3	.0310217	.0019026	.0272836	.0347598

	Mean	Std. Err.	[95% Conf.	Interval]
change1	.6176284	.0047568	.6082826	.6269741
change2	.3517885	.0046772	.342599	.3609779
change3	.0305832	.0015603	.0275177	.0336487

Tables give a clear message: the birth-place closeness with other agents does not affect the choice of agents. Anyway we must be careful trying to understand what went wrong. We could blame the model or we could blame the measure itself.

If we doubt of the model itself it means that the assumption of more meetings, more chances of becoming AT does not hold. In order to test this

assumption we ran an experiment, always in standard setting, on variable *contact* which is described in section 3. If the model works properly we should find that students who became AT have met more people. Let's take a look at the graph



The evidence in the graph is strong enough. Both the range and the mean of the number of meetings happened to students who became AT is much bigger with respect to the other group. Then we can consider our assumption as a truthful one, at least in the model.

If it is not model's fault it must be because of the measure we chose. Detecting the coordinate where an agent is born and measure a proxy of the closeness of other agents is not a proper way to find a determinant of student's choices. It turns out that the proximity of the the birthplace it is not strong enough to overcome the randomness of the movements. Another puzzling result about this measure is that AT students have the same e-distribution of IT students as can be seen in the following table

	Mean	Std. Err.	[95% Conf.	Interval]
ATe1	.6101563	.0031014	.6040628	.6162498
ATe2	.3607697	.0031943	.3544937	.3670457
ATe3	.029074	.0010486	.0270139	.0311341

This can be explained considering that the binary variable *e1* assumes value 1 even

if only one other applicants is close: the value of *e1*, as well as *e2*, does not change whether there are hundreds of people around you or only one.

Because of all previous analyses we can not rely also on measure of *d1,d2*. Observing results tables from a standard experiment in the second period we could conclude that admitted students from the previous period do not affect applicants choices in the second period. Anyway this would be a wrong conclusion, since the measure itself does not work properly. Hence we can not conclude anything about the role of already admitted students.

	Mean	Std. Err.	[95% Conf.	Interval]
ITd1	.0238572	.0016738	.0205687	.0271457
ITd2	.0527268	.002399	.0480135	.0574401
ITd3	.923416	.0030121	.917498	.9293339

	Mean	Std. Err.	[95% Conf.	Interval]
changed1	.0247757	.0014227	.0219804	.027571
changed2	.0503269	.0020643	.0462711	.0543827
changed3	.9248974	.0025495	.9198884	.9299064

4.8 A policy experiment

Nowadays, all around the world there has been a clear trend: selective universities are more and more interested in attracting high-achieving, low-income, students. This trend is made evident by the extremely generous scholarships offered by prestigious institutions such as Harvard or Stanford. Unfortunately, it seems that in real world, as in our model, some high achieving low income students still do not apply to top university. Why is it the case? As shown in the reference paper and in our model it is due to a lack of information.

How can information spread more quickly and reach all the applicants? So far, we have analysed how the results of the model change as some environmental characteristics change. For instance, let's consider *experiment 2*. It analyses how the percentage of students who change their behavior varies when the number of high-achieving students increases. Even though the result is an interesting one, there is no way for selective universities to practically exploit it. Indeed, in real world the number of high achieving students is given (or a policy to increase it can not be implemented by a university alone, because it would require a deep social intervention).

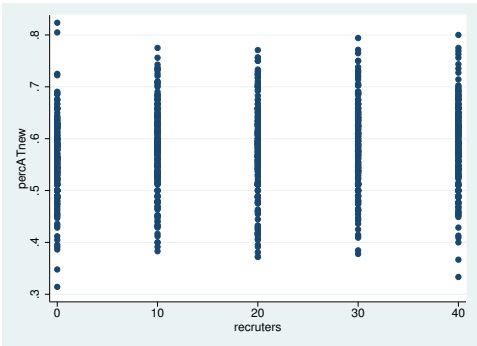
Thus, we need to build up a radically different experiment in order to model a policy that selective universities around the world (and the one in

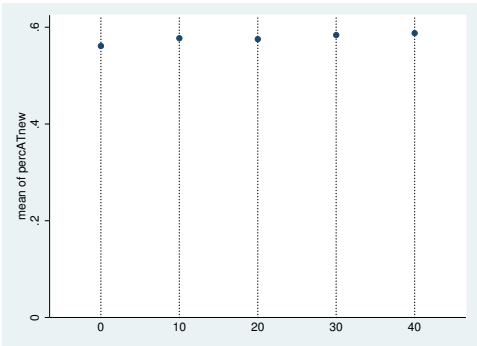
our model!) can exploit. So, what can a top university do to attract these missing students? As for a sport team who sends its recruiters all around the country to identify the most talented players, a university can send its *recruiters* to induce an increasing number of high achieving students to apply.

Is this policy effective? In order to check whether this policy is effective or not, we created a new agent set in the model: the *recruiters*. Clearly, some assumptions about recruiters have to be introduced:

- **Policy Assumption 1- Sensitivity of applicants to recruiters** Applicants are influenced by recruiters as strongly as by top university's students.
- **Policy Assumption 2- Recruiters move fast to search for high achieving students** Recruiters travel around the world faster than any other agent in order to meet the highest number of high achieving students.

How have these two assumptions been translated into the code? Recruiters as well as top university's students are typified by attribute *period* = 1, therefore we asked students to react in the same way when meeting every agent with *period* = 1. In this way, *Policy assumption 1* is satisfied. In order to get the second policy assumption satisfied we built up a new command *move-recruiters*: recruiters move randomly by 0.8, whereas all other agents' step is of 0.3. Now we are ready to answer the original question and to test whether the policy is effective.





We run 200 standard simulations for an increasing number of recruiters. The number of recruiters take values: 0, 10, 20, 30, 40. As it can be seen from the chart above the effect of recruiters is not strong at all. Low income students are still too scattered on the rural area and recruiters have only a limited

time to travel around and meeting them. This result confirms what Hoxby and Avery found: universities alone can not be very successful in attracting low income students who lives in rural area.