



Les sciences sociales en question : grandes controverses épistémologiques et méthodologiques

Compte rendu de la 67^e séance

Measuring racial discrimination: field experiments in the streets of Paris

November 6th 2023

Nonna Mayer (Sciences Po, CEE, CNRS) starts by saying how happy she is that she has good news concerning our colleague Fariba Adelkhah, who, after four and a half years, has just returned to France, free. Then she introduces the topic of today's seminar: field experiments and the measure of racial discrimination¹. Field experiments are commonly used to measure discrimination in non-interactive situations. A host of "correspondence studies" (testings) examine differences in callback rates to job applications according to the origin or religion of the applicant. But they rarely consider discrimination in everyday social interactions, despite the psychological distress such discrimination is known to generate. Martin Aranguren² (Sciences Po, CRIS) took on the challenge of investigating everyday discrimination in urban public places, on the

¹ Proceeding recounted by Justine Brisson and reviewed by the two presenters.

² Martin Aranguren has recently published : "Responses to the Islamic headscarf in everyday interactions depend on sex and locale: a field experiment in the metros of Brussels, Paris, and Vienna on helping and involvement behaviors", *PLoS ONE*, 2021 (with Francisco Madrisotti, Eser Dumas-Martin et al.); "Anti-Muslim behavior in everyday interaction: evidence from a field experiment in Paris", *Journal of Ethnic and Migration Studies*, 49(3), 2023: 770-794 (with Francesco Madrisotti & Eser Durmaz-Martins); Special issue (ed.) "Prévalence, mécanismes et conséquences cumulatives des discriminations en raison de l'origine", *Appartenances & Altérites*, 3, 2023.

basis of field experiments. He speaks about a large-scale field experiment that he recently conducted on racial discrimination in the streets of Paris, exploring the help people are ready to give to passers-by. His research is discussed by Morgane Laouenan³ (CNRS, Centre d'économie de la Sorbonne -Paris 1 – Sorbonne, IRD/LIEPP-Sciences Po), a specialist of racial and gender discrimination, and by Achim Edelmann⁴ (Sciences Po, Medialab), a specialist of the sociology of culture, social networks, and computational methods.

Martin Aranguren's intervention

Martin Aranguren thanks Nonna Mayer for organizing this seminar. He explains that he conducted a large-scale field experiment in the streets of Paris in 2021 to measure racial discrimination in terms of how people choose to help or not help passers-by. This is part of a broader agenda concerned with the explanation of macro-social (i.e., population-level) inequalities in mental health between groups that are stigmatized and groups that are not. Today's case illustrates the micro-social mechanism he is interested in, the occurrence of discriminatory behavior in a situation of social interaction. His research question is: To what extent does altruistic behavior depend on the target's social identity (e.g. the target's "race")? The measure of discrimination used here is observed differences in helping behavior.

Martin Aranguren recalls the accepted definition of altruistic behavior: acting out of consideration for someone's need. The person who helps will be called the "helper" and the person receiving the help will be called the "target". The helping situation under study occurs in a public space. Helper and target are both strangers to one another. The academic literature usually focuses on antecedents of helping behavior, treats helping as an indicator of prejudice or social capital, and considers its consequences for mental health. In the study he is presenting for this seminar, Martin focuses on only

³ Morgane Laouenan has recently published "Ethnic Discrimination on an Online Marketplace of Vacation Rentals", *American Economic Journal: Applied Economics*, 14 (1), 2022 (with Roland Rathelot); "Discriminations à l'embauche des personnes d'origine supposée maghrébine : quels enseignements d'une grande étude par testing ?", *Note IPP*, 76, 2021 (with Mirna Safi, Pierre Villedieu et al.) ; "Discrimination à l'embauche selon le sexe : les enseignements d'un testing de grande ampleur", *DARES Analyses*, 26 & *Note IPP*, 67, 2021.

⁴ Achim Edelmann has recently published "Computational Social Science and Sociology", *Annu Rev Sociol.* 2020, 46(1) : 61-81 (with Tom Wolff, Danielle Montagne, Christopher A. Bail) ; "Formalizing symbolic boundaries", *Poetics*, 68(1), 2018.

two types of factors: the characteristics of the target (race and social class) and of the situation (the cost of the form of help involved), as indicators of prejudice. He observes whether people differ in the rate to which they help the target, depending on his/her social class or race, and depending on the cost of helping. In this experiment, a confederate (in the role of the target) initiates an interaction with a randomly selected pedestrian (in the role of the helper). The experience was conducted between July and September, 2021, at four different places in the streets of Paris (place de la République, place de la Bataille de Stalingrad, place de Clichy, and in the vicinity of Saint-Lazare station, all close to busy metro/train stations) resulting in a total number of over 4,500 interactions. The targets differed according to their perceivable race and gender, and appeared in clothing connoting either a higher or lower social class status. At the places where the experiment was performed, pedestrians turned out to be mostly white and casually dressed. The diversity of the sample of pedestrians on the dimensions of race and class is therefore relatively limited.

The first question is: do randomly selected pedestrians (helpers) give more or less help depending on the target's race? The second one is: do they give more or less help depending on the target's social class? Martin's predictions were: 1) pedestrians (helpers) will offer less assistance when the tester (target) is Asian or Black (vs. White); 2) pedestrians will offer less assistance when the tester signals higher class status (vs. lower class status); 3) racial discrimination will increase when helping is more costly. The third prediction is the trickiest because one must define discrimination. Martin draws from the social psychology of discrimination, largely based on social identity theory. Discriminatory behavior will occur when the situation of social interaction is understood by the actor as an intergroup situation as opposed to an interpersonal situation. Martin only focuses on the third prediction in this talk, namely that racial discrimination will be greater when helping is more costly.

The helping situations took place in the streets of Paris, as mentioned before. The first example is an interaction between a tester and a randomly selected pedestrian of the same gender, the first mimicking a casual situation of social interaction. The procedure basically starts with an interruption from the tester ("Sorry to bother you...") to the pedestrian. At this point, the pedestrian can refuse to interact, or can accept the interaction. If he or she accepts, the tester asks: "I am looking for XXX street". Then

the pedestrian can choose to give directions, or not. If not, the tester insists, saying: "Because my phone has crashed. I can't find the street". The pedestrian can then accept to help and search the street on his/her own phone, or not. If not, he can help in another way (guessing or sending the tester elsewhere for help, for example), or not. It is obvious that searching on the phone is the costliest situation because it takes time, and it can be dangerous for the pedestrian (risk of theft, for example). One would expect discrimination to be stronger when it comes to searching on the phone.

As for the results of the experiment, Martin starts by describing the sample obtained. There were slightly more interactions with young people (18-25 or 30-49) than for older ones (50 and above). Only a very small number of pedestrians refused to interact. Most of them just gave the direction of the street. Searches on the phone represent half of the situations when help was given. Then Martin tests the proposed hypotheses. As a reminder, one was that "Helpers will offer less assistance when the tester is Asian or Black vs. White" and the other that "Racial discrimination will be greater when helping is more costly". Indeed, Black targets appeared to be discriminated against overall, but much more so when helping was more costly. There was no observable overall class discrimination. But combining the effects of the situation, gender, race, *and* class gives contrasting results. When helping is costly, Asian male testers were more discriminated against, Blacks appearing to be of higher status were less discriminated against than those in more casual clothing, and conversely White men appearing to be of higher class status were more discriminated against than those dressed more casually. When White men look more elegant, they are offered less help!

Martin considers that the experiment was successful overall. The results of the French experiment for Black targets replicate the results reached for Blacks in the United States, as well as for Asian men (but not women). In addition, the French experiment represents a successful methodological advance, introducing the variation in the cost of helping in the design of the experiment. The exploratory findings highlight the decisive impact of high-cost aid. As for the meaning of the class signal it seems to depend on the tester's apparent race. It means something different to look like a manager, or to dress more formally at any rate, depending on whether one is Black or White. The conclusion is that racism persists but under less obvious, surreptitious

forms, because of the social norms prohibiting or discouraging the overt expression of racism in democratic societies. It takes almost covert forms. In the experimental situation, introducing variations in the cost of help brings this subtle racism to light.

Morgane Laouenan's discussion

Morgane Laouenan thanks Martin Aranguren for his presentation on a topic she finds very interesting. She focuses on a comparison between “audit studies”, the generic name given to the type of field experiment Martin conducts,⁵ and correspondence (testing) studies. The main critique of audit studies is that the experimenter's behavior cannot be controlled. This is especially true for discrimination that occurs in the labor market, on which her research focuses. There are drawbacks and advantages in both methodologies – audit studies and correspondence studies. The two are both important. Audit studies are very expensive and produce less, although perhaps better, data.

She wants to know more about the selection process of the pedestrians. Were they randomly selected? She also wonders if traffic lights can influence the behavior of the helper. If a light turns red, then it means that the helper will have little time, less than a minute ahead of him, to cross. She is interested by the finding that there was less discrimination, sometimes none, against Asians. She is surprised by this, especially since the Covid pandemic. A lot of papers have shown that the pandemic increased discrimination against Chinese people in particular, who are seen by some as responsible for the propagation of the virus. Last, she wants to have more information on the underlying mechanisms behind this result.

Achim Edelmann's intervention

Achim Edelmann thanks Nonna Mayer for the invitation and Martin Aranguren for his presentation. He considers that Martin's work is essential because field experiments are key to knowing more about discrimination. His first remark concerns the theoretical

⁵ An audit study is the name given to this type of field experiment primarily used to test for discriminatory behavior in order to avoid social desirability bias.

framing. For him, it is baffling to start as Martin did with social identity theory. He wants to know more about this choice of theoretical framing. He also would like to hear more about the (antiracist) social norms Martin was referring to in the beginning of his presentation. How do social norms play a role in individual behavior in these cases since they involve an interaction between just two persons? No one is there to judge the behavior of the helper?

He also considers that red lights can influence the helper's behavior. He wants to know more about self-selection among the pedestrians. He considers that it would be interesting to ask those who refused to help when testers belong to visible minorities *why* they refused. Overall, he considers that the work is a fantastic project with incredible potential for future experiment.

Martin Aranguren's response

Martin Aranguren thanks the two discussants for their questions. He prefers the term "in-person studies" to "audit studies". In the field of economics, there have been many criticisms against in-person studies. Some of these criticisms are justified, but not all. Most apply to studies in which nobody is there to watch the way the experiment runs, or in which the participants of the study are aware of the hypotheses under examination, and even decide to work as confederates in the experiment because they are committed to the struggle against racism. This was typically the case for some studies commissioned by US public agencies in the 1960s -70s. These biases can be easily avoided with the routine and universally accepted safeguards that are used in social psychological studies on social interaction.

He then comes back to the problem of traffic lights. In selecting people to approach for help, testers selected the first person to arrive after the traffic light turned red. This solves the problem of self-selection, because pedestrians generally don't choose the time at which they come up to a traffic light. They just happen to arrive at the pedestrian crossing at some point, and at this particular juncture the light may be green or red. The large and sustained volume of motor traffic on the road at rush hour ensures that every pedestrian, no matter their motivation or physical condition, will stop when the light turns red.

Martin had read about the rise in anti-Asian feelings during the Covid period and this is precisely why he included a group of Asian testers in the experiment. However, the study showed that these antipathies don't extend to helping behavior, at least for women.

Emilien Houard Vial's questions

Emilien Houard Vial (CEE, Sciences Po) has several questions. He wonders if it is true that women are less likely to discriminate, as some literature seems to show. The second question is about the choice of the places for the experiment. There would have been a more diverse population if the locales chosen contrasted the 16th or 17th arrondissement of Paris with the neighborhood of Barbès in the 18th for example. He also wants to know more about the research ethics, for instance if confederates were told of the potential risks such an experience might induce. They could have been confronted with brutal forms of racism for example.

Martin Aranguren's response

Martin states that there is no evidence whatsoever showing that women discriminate less. The rates are very similar. As for the places chosen, the experiment included sites that differ on a number of obvious dimensions. However, the main reason why they chose the places they did was a conjunction of two things. The first is having a sufficiently busy road so that pedestrians would not cross during the red lights. In a busy street, they do have to wait in order not to take risks. The other requirement was a densely populated place, in order not to lose too much time between each interaction.

He confirms that there are research ethics principles to follow when doing this sort of experiment. This includes consideration of the risks involved, and the conclusion that these risks are overridden by the potential benefits of the study. Confederates were aware of the risks they would be exposed to. They just didn't know the exact research hypothesis. Martin only showed them the design and hypothesis of the study once the data collection was complete. Obviously, participants saw that the people asking for

help came from different racial groups and different genders, but they did not know why. They also knew that the research would be published afterwards.

Gabriel Feltran's question

Gabriel Feltran (CEE, Sciences Po) wonders why there were not any blond or blue-eyed people among the confederates. Why did Martin not test this? Could there be something like positive discrimination? He takes the example of Brazilian people. In Brazil, many Brazilians would consider themselves as White, while in Europe they would be deemed persons of color. The definition of White and Black may differ from one country to another. He also wonders if Martin took advantage of the expertise acquired by his team. Did they discuss their experience after?

Martin Aranguren's response

Martin Aranguren completely agrees with the fact that there is a totally local understanding of race. Being White in France is not the same as being White in Brazil or in the United States. But in this case, he wasn't much concerned about the specific meaning of race. He just focused on people that most people would consider as White (or Black or Asian) in France. He just used common sense perception. He agrees that there are degrees of "Whiteness", but he just used prototypes. It would be interesting to explore nuances in a study on Whiteness but in this specific study this wasn't much of an interest. As for the post-experiment state of mind of testers, it was very interesting to observe. Most of the participants were acting in a routine manner. They became quite insensitive to what was happening after a few hours of experimental practice. They did their job very well and became better at acting as time passed, but they didn't pay much attention to the pedestrians' reactions. It was the opposite at the beginning, when they were more hesitant and sensitive to the pedestrians' responses. The process that saw them become more and more proficient in their own acting appears to be correlated to an increased insensitivity or indifference to what the pedestrians might do on their side. They had less and less to tell as the days went by.

Questions from Lena Schorlemer and another participant (not identified on the tape)

Lena Schorlemer (CEE, Sciences Po) has two questions. How did the testers deal with respondents who were not fluent in French or who refused to help the testers because of their language gap, i.e. were their answers treated as a refusal to help (possibly distorting the results) or were they excluded from the observations? On what basis were the test sites selected?

Another participant asks: since the testers were in the same street on multiple consecutive days, perhaps they could have been recognized by pedestrians who would also pass through this street daily. This could put them on their guard, causing them not to answer?

Martin Aranguren's response

Regarding the sampling, Martin agrees that it is theoretically possible that people might have recognized some of the confederates. They stood there for hours in the same place; it is even possible that some people could have been approached twice. Fortunately, this did not happen. None of the locations selected was a square, so people don't really stand for long at the traffic lights; they are on the move. Also, there was a very specific and complex schedule that made duplications unlikely. The same confederates did not intervene in the same place for four days. And when they came back to it, their outfit was different. So, the probability that they might have been recognized, while certainly not zero, is very low.

As for pedestrians not fluent in French, they were not included in the sample. If a pedestrian displayed lack of proficiency in French after being approached by the tester, the attempt at interaction was immediately terminated and excluded from the sample.

Jean-Baptiste Chambon's questions

Jean-Baptiste Chambon (Sciences Po, CEE) wonders why very middle-class looking targets (the confederates) were chosen when they were supposed to illustrate a lower-class status. He considers that their clothing is not representative of all young adults.

He also wonders whether the participants were trained to “perform” like the people they were supposed to represent, for example as a poor young Black man. Last, he questions Martin about his conception of race. It seems to him he considers race as an ethno-phenomenon, essentializing groups, and is actually missing the point of race as a social construction, an evolutive and dynamic process.

Martin Aranguren’s response

For him, it doesn’t matter if the clothing that conveys a relatively lower status in the experiment connotes a middle-class status. All that matters is that the status being communicated by that clothing is lower than the status signaled by the other clothing. They just wanted to examine the potential difference in treatment arising from a difference in status as conveyed by clothing. Regarding the second question, he explains that he didn’t ask the participants to perform a race or a gender. He just wanted to measure the difference, for instance, in reactions to people belonging to two different racial groups, acting the same. They did not want participants to “play Black”. Racial categorizations are to some extent independent from the performance of a racial stereotype. In this experiment, pedestrians were assumed to categorize testers as White, Black, or Asian on the basis of the tester’s appearance.

Eva Bossuyt’s question

Eva Bossuyt (CEE, Sciences Po) wonders if Martin and his team conducted in-depth interviews with the confederates, especially if the latter personally experienced discrimination in their everyday life. She also wonders if the age of the pedestrian could not play a role in this experience. For example, it might be less costly for an elderly and retired person to take time to help someone than for someone working and having very little time during the day. In contrast, younger people are more likely to be at ease with using mobile phones than the elderly, which would similarly reduce their cost of helping. Eva therefore asks whether results were broken down by age group during the exploratory study and if age was significant for explaining differences in the decision to help.

Martin Aranguren's response

The effect of age was minimal. It didn't affect the study much. The older the pedestrians the less inclined they were to use their phone to find the street, but the effect of age was the same whatever the race of the tester.

Questions by George E. Marcus

George Marcus (emeritus professor of political science and political psychology, Williams College) expresses that he thinks more field experiments like the one presented by Martin Aranguren should be completed. He then poses a series of questions and remarks.

One, the discrepancy of age between the testers, who look young (on the PowerPoint presentation) and the helpers should be taken into consideration. People tend to trust someone of their same age group more than others. Two, another factor that could matter is body mass; frailty can make people feel more vulnerable in social interactions. Three, everything depends on the context where the experimentation takes place. In a touristic context one may be on the lookout, afraid of scams; perceived vulnerability can distort the interaction between tester and helper. George adds that Paris neighborhoods are like villages, with their specific codes and rules; someone who does not respect them can be resented as a stranger in the environment. Social norms matter in ordinary social interaction as shown by Erving Goffman.

George Marcus brings attention to the fact that the purpose of discrimination is not the same in the ordinary situation of asking for help studied in the experiment as it is in the labor market or the housing market, where the issues at stake are far more serious. He also stresses the necessity to take into consideration the way the human brain functions. The moment one sees somebody, they unconsciously register a lot of things about the person, in terms of age, race, gender, and social status. And the image that results is not necessarily true as shown by the cognitive scientist Donald Hoffmann in *The case against reality* (2016).

Nonna Mayer ends the discussion with a warm thanks to everyone, inviting all participants to continue the discussion at a café.