DISCRIMINATION IN THE SHARING ECONOMY: EVIDENCE FROM A HUNGARIAN FIELD EXPERIMENT ON THE TOPIC OF RIDESHARING

Bori Simonovits, Iryna Shvets, Hannah Taylor[1]

ABSTRACT *This paper describes small-scale experimental research (total N=160) implemented in Hungary in summer 2017 by TÁRKI, the Hungarian Social Research Centre. The research was based on a controlled field experimental design and examined the chances of ethnically different testers being offered rides on the most well-known international ridesharing platform. The creation of eight experimental profiles enabled us to test the interaction between two experimental stimuli: ethnic/racial background and gender. The present paper provides evidence of ethnic/racial disproportionality in offers of rides to the different testers by multi-ethnic drivers. A large-size effect was detected in the case of the Arabic male tester and medium size effect in the case of the Chinese male tester. In the cases of a Russian male profile and female testers of all nationalities no evidence of racial or ethnic discrimination was detected compared to our Dutch benchmark profile. Although the pattern is not clear, an interaction effect was present between gender and nationality; as a consequence of this interaction, the Arabic male tester had far less chance of being offered a ride compared to any other combination of gender and nationality.*

KEYWORDS: *discrimination, sharing economy, ridesharing, controlled field experiment, Hungary*

1 Bori Simonovits, PhD. is a Senior Researcher at TÁRKI, Tárki Social Research Institute Inc, e-mail: simonovits@tarki.hu. Iryna Shvets, M.A. is a Ph.D. candidate in Political Science at Opole University, e-mail: ishvets@uni.opole.pl. Hannah Taylor is a fourth-year student at New York University Abu Dhabi, e-mail: hct245@nyu.edu.

INTRODUCTION

Over the centuries people have searched for and engaged in different ways of resource sharing to save costs. For instance, the first car-sharing service was launched in 1948 by a small housing cooperative in Zurich and later, in the 1980s, car-sharing services owned by small community-based cooperatives became very popular, especially in Northern Europe (Shaheen, Sperling, & Wagner, 1999). Modern ridesharing services are usually classified as part of the sharing economy or as a form of collaborative consumption, described by Hamari, Sjöklint and Ukkonen (2016) as "peer-to-peer-based activity of obtaining, giving, or sharing the access to goods and services, coordinated through community-based online services."

The online component of modern sharing economies is vital for their existence. The internet fosters their development – minimizing the cost of transactions and communications and making organizationally complex services easily accessible around the world. Belk (2014) defines the sharing economy as having two parts: collaborative consumption and sharing itself. Access to information, services, and products is heightened as users need only the following resources to participate: a browser, internet access, and a government that allows access to most or all web content. Both sharing and collaborative consumption rely on temporary access rather than ownership models for consumer goods and services and employ the internet in the form of "Web 2.0" which allows users to contribute content and connect. This contrasts with the one-directional provision of information to customers, the main feature of the internet in the past.

Nowadays, so-called sharing platforms are operating in the most important areas of the economy such as accommodation, transportation, rental, retail, and logistics, and therefore affect the entire economy. The global rise of sharing economy platforms such as Uber and Airbnb has increased concern about their impact and divided the public. While citizens and taxi drivers alike protest on the streets, Airbnb (2017) and Uber (MacDonald, 2014) have published reports in which they claim to have positive impacts on the economies of cities in the US and in Europe.

While the debate about the proper labelling and understanding of the roles of such platforms is still ongoing (Codagnone & Martens, 2016), the fact that they represent a unique opportunity for social research is evident. For example, studies may examine why people participate in sharing economies (Hamari, Sjöklint, & Ukkonen, 2016) and how government policy should treat them in relation to other competing services (Koopman, Mitchell, & Thierer, 2015). In particular, Codagnone and Martens (2016) report that building trust between both sides of the sharing economy market has been a challenge even for large

enterprises. Ridesharing platforms provide one example of the sharing economy where the success of the shared activity is based on mutual trust between driver and passenger. Many platforms have made an attempt to quantify the level of trust by introducing review scores and ratings, but this does not guarantee that the provider will accept the offer of a possible customer in the presence of alternative considerations like xenophobia and racial or ethnic intolerance.

Empirical qualitative research has been conducted in this field in Hungary by Dén-Nagy and Király (2014). There is empirical evidence that discrimination occurs in the Hungarian labor market based on controlled field experiments carried out via phone calls, in-person, and the sending of CVs by email to job vacancies (Pálosi, Sik, & Simonovits, 2007; Sik & Simonovits, 2008). Nonetheless, measuring discrimination in the sharing economy has never been systematically studied in Hungary. Our paper therefore describes research designed to measure the discrimination of drivers against different racial/ethnic groups through the application of a controlled field experiment that involved a specific international ridesharing service that operates in Hungary.[2]

In the present paper we focus our attention on the following research questions:

1. Is there evidence of ethnic/racial discrimination in the sharing economy; more specifically, in an international ridesharing service operating in Hungary?
2. Which ethnic group(s) are the most likely and which the least likely to be offered a ride from the following: Dutch, Russian, Chinese or Arabic?
3. Does gender matter in this context?

The paper is structured in the following way: after an introduction (Section 1), the theoretical background of the sharing economy and previous research in the field is discussed (Section 2). This is followed by the formulation of the research hypotheses (Section 3). In Section 4 we present our research methodology then discuss our results (Section 5). Finally, conclusions are drawn, including a discussion of limitations (Section 6).

THEORETICAL BACKGROUND
AND PREVIOUS RESEARCH IN THE FIELD

The role of trust in the sharing economy

The sharing economy necessitates trust from users in order for their successful participation, as risks within the newest form of P2P sharing economy (which not only involve direct sale or simple sharing but rather mutual participation) go beyond the monetary. There are risks inherent to sharing one's home with someone, and services are intangible and non-verifiable rather than products to be sold. To create trust in response to these risks, users tend to provide more robust information on the newer forms of P2P sharing economy profiles. For example, on Airbnb, directly next to photos of the accommodation, are photos of the host.

Sztompka's sociological framework of trust (2003) considers trust as a necessary response in the face of unknown future circumstances and outcomes. Online sharing economies, which have a certain level of anonymity and leave room for deceit, create such unknown circumstances. Trust is identified on three levels; (1) the individual level, or as a personality trait, (2) at the level of interactions, as in the quality of a specific relationship, and (3) at a cultural level (Sztompka, 2003). The user's cultural background dictates their ability to trust online by teaching individuals whether to have high or low trust, what to expect of individuals in different roles, and their level of communicability, which is shown to create and foster trust (Dén-Nagy and Király, 2014). Dén-Nagy and Király posit that those who join sharing networks are likely have an above-average propensity to trust at the personality level. The researchers found that the range of associated risks were generally thought to be small, and centered more on risks to emotional wellbeing such as the risk of encountering an awkward situation. Risks of safety to person and property were mentioned with less frequency.

Homophily was found to relate to age and personal interests, while cultural background fell into the realm of preference. Even outside of the online space, users credit CouchSurfing with offering the sense of a trusting community that is lacking in their lives. In our research, homophily is examined as it relates to ingroup and outgroup identity in terms of race and nationality as well as age.

Ert and colleagues (2016) examined the role of photos of hosts on perceptions of trustworthiness of guests on Airbnb, a home-sharing rental platform that has flourished within the bounds of the sharing economy. The researchers empirically analysed data available on the platform as well as conducted a controlled experiment in which participants were recruited on Amazon's MTurk and paid to judge the attractiveness and the trustworthiness of photos, and to

state whether they were likely to stay in various lodgings depicted in the photos. Ert et al (2016) found that the effects of the photos of the host – which serve to establish visual-based trust – had more influence on trust than reviews left by other users. In general, there is a low variance between Airbnb review scores, and an examination of five large European cities revealed the average rating to be between 4.5 and 5 stars. Ert et al. also found that hosts that are perceived by users as more trustworthy charge more for their accommodation than their less-trustworthy counterparts. Female hosts were preferred over male hosts. The interplay between trust and reputation merits examination, as trust can be fostered by reputation in the form of ratings, but is mediated by other factors such as visual-based trust. Additionally, when starting using a sharing economy platform, users do not have a reputation on which they can build trust, and a photo serves as the first mechanism.

Beyond personalized photos, feedback mechanisms also play a crucial role in creating trust on collaborative consumption platforms. In their 2002 paper, Resnick and Zeckhauser examined eBay, a virtual commerce site which combines user-to-user and business-to-user sales. The site collects comments about users after each transaction. Examining transactional data from 1999, the researchers found that the online trust community was facilitated by readiness to provide comments, that reputational profiles predicted future performance, and that there was a strong correlation between the feedback provided by the buyer and the seller, suggesting reciprocity in interaction. Both buyers and sellers reviewed their partners over 50% of the time, and 99% of all reviews were positive, more than would be expected. Additionally, good reviews predicted more sales in the future, but did not increase the profitability of each sale.

The importance of trust-building in ridesharing platforms was also outlined by the joint research between BlaBlaCar and NYU professor Arun Sundararajan. The Europe-wide empirical research effort by BlaBlaCar and Sundararajan – based on responses from more than 18,000 BlaBlaCar members across 11 countries – concluded that: "We are at the dawn of an incredibly promising era. Together we are entering the trust age," (Mazzella and Sundararajan, 2016, p. 39). Their data revealed that the role of full profiles, including photos, reviews, and verification, as well as the brand effect of BlaBlaCar, were crucial for creating digital trust, making riders and drivers who had never met trust each other almost as much as family members or friends.

Knowing the importance of reviews as a mirror of and support for trust in the sharing economy, we incorporated the feedback of clients – in terms of reviews and ratings – from drivers we interacted with into our analysis. Similarly to eBay, on the ridesharing service we analyzed, reviews tend to be overwhelmingly positive.

Discrimination in the sharing economy

Beyond the role of trust and risk in the sharing economy, the role of discrimination merits exploration. Discrimination can be considered to fall into one of two categories, depending on the root cause. Arrow distinguished between statistical and taste-based discrimination, an approach which is underlined by rational choice theory (1998). Rational choice theory asserts that beliefs are based on evidence, and those which are contradicted do not survive. This gives rise to statistical discrimination, in which measurable differences in groups lead to discriminatory attitudes between them. To assume that a discriminatory attitude is statistical, one must assume that the discriminator uses all sufficient available data when deciding. Taste-based discrimination (Becker 1971), on the other hand, is based on personal preference rather than group difference that can be statistically proven.

After analyzing Airbnb listings in New York City, Edelman and Luca (2014) found that compared to white Airbnb hosts, black Airbnb hosts advertised their accommodation for less. They found that for equivalent rentals, black hosts tended to advertise their accommodation for 12% less, and the authors assert that the reduction in rents on the part of black hosts is a response to discrimination that makes their listings less favorable. While built-in mechanisms such as host photos are intended to create more trust between users, they may have the unwanted effect of leading to racial discrimination.

In a similar vein, Cui et al (2016) created fictitious guest accounts on Airbnb and sent requests for accommodation to 1,508 hosts. The researchers found that, when compared with requests from white-sounding names, requests from guests whose names signalled they were African American were 19.2% less likely to be accepted. In line with the emphasis placed on reviews in the associated literature, reviews significantly reduced the likelihood of rejection. The authors claim that this is an example of statistical discrimination, in which first judgments are made based on the appraisal of the racial group, but evaluators are amenable to changing their judgment based on new information.

In their 2016 research, Ge and colleagues had passengers hail approximately 1,500 rides in Boston and Seattle and recorded the performance metrics of the Uber drivers. They found discrimination based on race in that waiting times for African American passengers were up to 35% longer in Seattle. In both cities, rides were more frequently cancelled when passengers had African-American sounding names. In low-density areas, ride cancellation for males using African-American sounding names was more than three times more likely. Race was not the only factor affecting treatment of passengers; there was also evidence that drivers took female passengers for longer and more expensive rides.

Finally, we devote special attention to the interaction effect of gender and race on discrimination. Elaine Shoben (1981) explored interaction effects in the area of labour market discrimination. The author explains that, depending on their personal characteristics, individuals can experience "double discrimination" if they belong to two adversely affected groups. Such discrimination can negatively affect potential employees on the basis of employers' practices regarding requirements for the job. When the interaction effects of race and sex (for example, applying to black women) affect workers' chances of employment in more subtle ways, this is what Shoben terms "compound discrimination." The presence of such discrimination, which does not exist solely because of race nor gender but instead due to the interaction of these factors, and manifests itself through employers' prejudices, is an issue worthy of examination in the shared economy.

Massey and Lundy's (2001) experimental study relied on the fact that racial discrimination can occur without personal contact, and that race can be inferred from communication over the phone. The researchers compared housing discrimination towards male and female speakers of Black English Vernacular, Black Accented English, and White Middle-Class English. Telephone calls with prospective landlords revealed that blacks generally experience less access to housing than whites, lower-class blacks (who speak Black English Vernacular) experience less access to housing than middle-class blacks (who speak Black Accented English), and that women have less access to housing than men. Race, gender and class interacted in a non-additive way in affecting housing access; i.e. poor black females experience the least access to rental housing.

RESEARCH HYPOTHESES

Scholars argue that both racial (see meta-analysis provided by Saucier et al. 2005) and gender differences (Eagly and Crowley, 1986) matter in prosocial behaviour. Focusing on the working mechanisms of the sharing economy, rational choice and statistical discrimination seem to play an important role according to recent research by Edelman and Luca (2014). Therefore, we considered it meaningful to formulate hypotheses concerning the effect of the gender and ethnic/racial characteristics of our testers as the most important predictors for discrimination. We were also interested in the effect of age on positive responses, which could be due to trust caused by homophily, but our research design did not allow us to manipulate the effect of age as we only used young testers. Therefore, we could only examine homophily as it related to the range of age of the drivers in relation

to the young tester profiles, lessening our ability to draw conclusions about the effect of age difference and trust.

We formulated several sub-hypotheses about the testers' race and gender, as well as on their interaction, to test the effect of trust, discrimination and homophily:

1. Drivers (mostly from Hungary and neighboring countries) will differentiate between the ingroup and the outgroup in terms of racial and national identity. As a consequence, drivers will be more inclined to automatically trust and offer seats in their cars to those who 'look European'; i.e., who are racially Caucasian.
2. There is distrust towards the outgroup, therefore those who have a different appearance (Chinese and Arab profiles) have a significantly lower chance of getting a seat.
3. Due to the current political climate, there is heightened distrust towards Arab people, leading to more rejections and less responses to Arab profiles.
4. There are higher response and approval rates for women than for men due to the tendency toward trust.
5. Furthermore, we suppose that there is a two-way interaction[3] between gender and perceived race, thus we hypothesize that Arabic males will be least trusted among all profiles.
6. There will be more positive responses when there is homophily between driver and passenger in terms of age.

As we did not use a control group in our research, we treat the Dutch testers as a benchmark category when calculating t-tests and odds ratios. We treat Chinese, Russian and Arabic profiles as the experimental variables.[4]

---

3 We understand interaction in the following way: Two variables interact if a particular combination of variables (in this case, being Arabic and being male) leads to results that would not be anticipated on the basis of the main effects of those variables. The main effect of an independent variable (here, gender and nationality) is the effect of the variable averaged over all levels of other variables in the experiment.

4 Even though we formulated certain hypotheses on the likelihood of positive feedback to experimental profiles (Arabic, Chinese and Russian profiles) by drivers, we could not test these because both the gender and age distribution of drivers were skewed to a large extent. This prevented us from testing the previously defined hypothesis about the drivers.

EMPIRICAL STRATEGY

Controlled field experiment design - methodology

Our empirical research relies on a *controlled field experiment*. In these studies, the controlled experiments are realized by two or more individuals (auditors or testers) matched for all relevant personal characteristics other than the one that is presumed to lead to any kind of discrimination. This area of testing can include specific areas of social life where discrimination may occur, such as the labor market (testers apply for a job), housing (testers apply for a housing unit as potential tenants or customers), the services sector, and the sharing economy (testers start to negotiate for a good or service). The treatment they receive and the results they achieve are closely observed, documented, and analysed to determine if the outcomes reveal patterns of differential treatment on the basis of the traits defined in anti-discrimination laws; most importantly, ethnicity, racial- or national origin, gender, or age (see more about the methodology and its implementation by Gerber and Green, 2012).

The method is widely used in empirical social research to test human behavior, such as the existence of pro-social or discriminatory activity. In our empirical research, we focused our attention on examining a specific and well-known international ride-sharing platform that operates in Hungary.

We must emphasize that the research presented here is a pilot study based on a controlled experimental design of only 160 observations. Therefore, results should be interpreted cautiously. Moreover, the chance of making Type II errors[5] while drawing conclusions is relatively high due to the low number of observations in the different tester's profiles.

As we wanted to test the effect of the racial background of potential passengers, as well as nationality (differentiating four nationalities) and gender in our field experiment, we used eight different photos and names to show these two characteristics on the ridesharing platform. Additionally, if drivers looked at the profile bios, nationality was shared. Control variables included age, social status (all testers were said to be university students) and language use (English, but not as a native language). We did not include a Hungarian passenger as a control tester as we wanted to keep language use constant, and with a Hungarian-named passenger drivers would have been able to communicate in Hungarian, which could have confounded our research. Therefore, we decided to use a non-Hungarian European tester as a benchmark; namely the Dutch profiles. This

---

5 A type-II error means accepting a null hypothesis (technically, failing to reject a null hypothesis) when the null hypothesis is false.

was due to the assumption that Dutch testers would be received neutrally or positively by drivers in the region. The next section shares how the research was carried out.

Implementation of the research design and fieldwork

The experimental research was implemented in Hungary in the summer of 2017 by two interns at TÁRKI under the supervision of a senior researcher (the three authors of the present paper). In line with experimental methodology, we have made as much effort as possible to strictly control all the relevant variables of the study.

   We created eight profiles on the ridesharing platform, representing four racial/ethnic groups (Table 1).

Table 1. Characteristics of testers (by nationality, gender, name and country of origin)

Nationality	Gender	Name	Country of origin
Arabic	female	Fatima	Lebanon
	male	Hassan	Saudi Arabia
Russian	female	Natasha	Russia
	male	Vladimir	Russia
Dutch	female	Julia	Netherlands
	male	Luuk	Netherlands
Chinese	female	Li	China
	male	Chen	China

The profile data also included passengers' year of birth. To control for age, testers were said to be between 25 and 28 years old (inclusive). Profile names and surnames were assigned based on a list of the most common names by country, available online.[6] Passengers' profile photos were also chosen to represent the relevant racial and ethnic background, while keeping other factors equal (most importantly, displaying similar levels of attractiveness). We also paid attention to characteristics of the photos, such as the direction the subject was facing and a lack of visible accessories. Faces were clearly observable and not covered by hair or anything else. Finally, all images represented smiling people with their teeth exposed. In addition, every profile was "verified" by both email and mobile phone.

---

6 http://www.studentsoftheworld.info/penpals/stats.php3?Pays=CHN

We also designed the text that appeared in the short biography of each user of the ridesharing service. The testers' texts looked alike, including a salutation and information that the passenger was a student based in Budapest. The rationale for identifying the testers as students was that students have the same and relatively high social status. Biographies also included the passenger's country of origin and a closing statement stating their love of travelling. For instance, the Dutch student's bio stated: "Hello! I am a student from Holland and I do my studies in Budapest now. Love to explore new places and travel around."

Our sampling strategy included selecting all drivers who were travelling to our predetermined destinations (See Table A1 in the Appendix), combined with the characteristics listed below. In general, there were two groups of directions for ridesharing travel; namely, rides starting from Budapest and leading to a destination inside of Hungary, and those which also started in Budapest but had a destination point in a neighboring country. Keeping in mind that the ridesharing service made it possible to automate the process – allowing for instant booking – we chose only drivers whose passenger acceptance settings were "non-instant approval." This ensured that drivers decided by themselves whether to pick up a passenger.

We further restricted our sampling to those drivers who did not state an exact meeting point or arrival point, and asked them for the missing information. In this way, we made sure that our message appeared reasonable to drivers and would never necessitate paying to book a place. Additionally, requests for rides were sent two days in advance of the time of travel, primarily in the morning hours between 10 and 12 a.m. This timing was implemented to give drivers time to consider the requests and respond so that the testers could rescind their requests, if positive. It is worth mentioning that the messages sent from Hasan's profile were sent in the evening, adding to the limitations of our study.

Our team sent 160 requests for rides, with 20 requests from each passenger to allow for statistical data analysis. We ensured that none of the drivers received the request for a ride two or more times. Our research design included soliciting rides for both domestic and international journeys from each profile.

Note on Research Ethics

In the case of field experiments, there are ethical issues related to informed consent; as in our case the experimental subjects (i.e. the drivers) could not be asked to cooperate beforehand, special care was taken to adhere to ethical standards of anonymity and confidentiality in line with the guidelines defined by TÁRKI's (Hungarian Social Research Institute) data-gathering department.

We anonymized data to protect the identity of drivers when collecting specific information about their interactions, as well as information about their profiles. Furthermore, we protected drivers from loss of income or undue stress by making sure we responded to them promptly after they accepted a request. We assume that due to our careful process of withdrawal, our subjects (the drivers) did not experience more than a minimum risk compared to a real-life situation (see more about the criteria of minimal risk in the Guidelines defined by the Journal of Experimental Political Science in 2014). We anonymized the specific aspects of the website that would give away its identity, as we did not want to expose the presence of discrimination on any particular ridesharing platform.[7]

RESULTS

For the descriptive portion of our analysis, we first briefly describe the composition of the drivers included in our sample as well as the characteristics of the ride postings and interactions that take place on the site which are shown by left-behind traces such as ratings and reviews. This is followed by an analysis of the different treatments received by the testers.

Drivers' reactions to passenger's requests

Observations came from interactions with 160 distinct drivers in relation to eight profiles. Of the drivers contacted, 62% were Hungarian and 27% came from neighboring countries – as was explicitly stated in their profiles or indicated by the language that their profiles were in. An additional 11% had national identities that were not clearly discernible. Neighboring countries included Serbia, Croatia, Slovakia and Romania. The self-reported ages for drivers varied between 18 years of age and 100 years of age. The average age was 35.3 years of age, about 10 years older than the ages of our tester profiles.

   On other sharing economy sites, especially non-monetary sites like CouchSurfing, bios and photos of service providers and recipients are the norm (Resnick and Zeckhauser, 2002). On the ridesharing site we analyzed, only

---

7 The anonymised database is stored in TÁRKI's data depository. Further information about the database is available in Hungarian at: http://www.tarki.hu/cgi-bin/katalogus/tarkifo_hun. pl?sorszam=TDATA-i30.
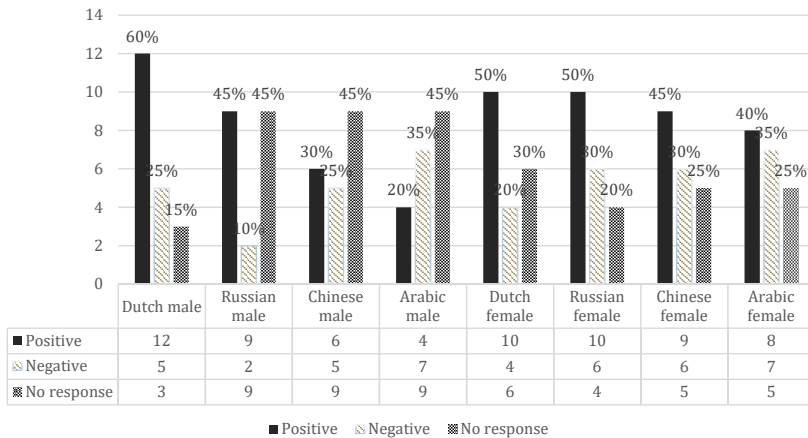
50% of drivers had written a bio for themselves, while 50% left their bio blank. Additionally, 15% of drivers did not upload a photo of themselves for their bio.

As far as ratings, experience and other relevant characteristics of the ride are concerned – consistent with the literature about feedback mechanisms in the sharing economy – reviews were overwhelmingly positive, and the average rating that drivers held was 4.7 stars, with a standard deviation of 0.47.

The main focus of our study was to make a comparative analysis of drivers' responses to our passenger requests. In particular, drivers' responses were categorized into three groups: positive, negative, and no response. A positive answer indicated the driver's agreement to provide a ride.[8] In contrast, drivers' responses were classified as negative when the driver refused to give a ride for any reason.[9]

In line with our expectations, the Dutch male profile (Luuk) received the most positive responses (60%) and the least amount of positive responses were recorded for the Arabic male passenger (20%), for whom 45% of requests remained unanswered. The Russian male received 45% and the Chinese male profile received 30% positive answers. Consequently, the distribution of responses for female passengers revealed that Dutch and Russian profiles received the same amount of positive answers (50%). In turn, the Chinese female profile recorded 45% positive answers while the same indicator for Arabic female was 40% (see Figure 1).

Figure 1. *Drivers' responses (% and count, categorized by prospective passenger)*

|	Dutch male	Russian male	Chinese male	Arabic male	Dutch female	Russian female	Chinese female	Arabic female
■ Positive	12	9	6	4	10	10	9	8
▨ Negative	5	2	5	7	4	6	6	7
▧ No response	3	9	9	9	6	4	5	5

■ Positive   ▨ Negative   ▧ No response

8 An examples of a positive answer: "Sure I can take you!"

9 An example of a negative answer: "Hello, sorry but I have not free place."

Another calculation was performed to establish the time that elapsed between the passenger's request and the driver's response (the mean was calculated for each passenger based on the time that elapsed after each request; see Figure 2).[10] The lowest mean time occurred with the Dutch male passenger (Luuk) at 5 hours and 9 minutes. The longest mean time span occurred for the Russian female passenger (Natasha) at 16 hours and 16 minutes. More statistical details about this measure can be found in the Appendices (Table A1).

Figure 2. *Mean response time between passenger's request and driver's response (hours and minutes)*
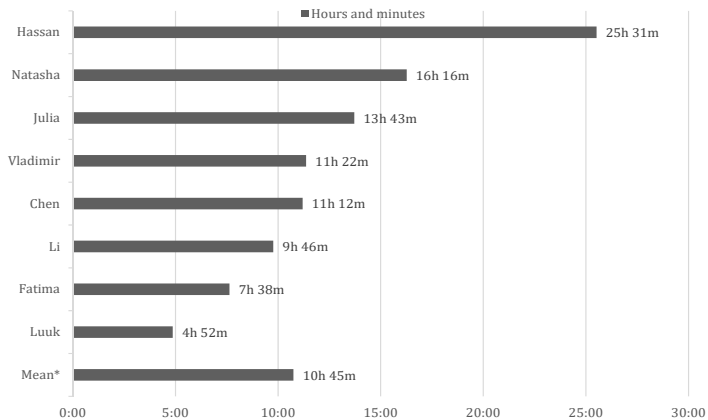
Treatment effect: difference in means and effect sizes

As already mentioned, in our experimental study we did not use untreated units, therefore the Dutch testers were used as "benchmark" profiles, meaning that Luuk and Julia functioned as control variables in the calculation of the treatment effect. The *standardized mean difference (based on Cohen's d)* of the experimental and the control group (summarized in Table 2) is the primary indicator for measuring discrimination. Later on, odds ratios for being treated favorably are also shown in the logistic regression models presented in Table 4.

---

10 The overall mean for passengers' response times was calculated excluding the data for the Arabic male passenger as the requests from this profile were sent in the evening while all others were sent in the morning. Consequently, this makes it harder to interpret the timing data for the Arabic male tester.

Comparing the profiles, it is clear that the greatest difference was found in case of the Arabic male tester compared to the Dutch male tester based on the mean values. While Hassan got positive feedback in two out of 10 instances, Luuk got six out of 10 offers from drivers. The effect size in the case of Hassan, the Arabic male, is evaluated as *large*, based on Cohen's d (d=0.87). A *Medium* effect was found in case of Li, the Chinese male (Cohen's d=0.62), while a *small* effect size was found in case of Vladimir, the Russian male tester (Cohen's d=0.29), in line with our hypothesis. As far as female testers are concerned, effect sizes are not noteworthy except for t in the case of Fatima, the Arabic female tester, for whom we found a small effect.

Table 2. *Characteristics of treatment effect on nationality: Mean of positive outcome, Standard Deviation, Cohen's d, by experimental tester vs. benchmark tester (N=20 for all profiles)*

	Mean (0-negative and 1-positive outcome)	Std Deviation	**Effect size (Cohen's d) Standardized mean difference**	Evaluation of effect*
Male **benchmark profile** (Luuk)	0.60	0.502	-	-
Male **experimental profiles**				
Vladimir	0.45	0.510	0.29	small
Li	0.30	0.470	0.62	medium
Hassan	0.20	0.410	0.87	high
Female benchmark profile (Julia)	0.50	0.513	-	-
Female experimental profiles				
Natasha	0.50	0.513	0.00	zero
Chen	0.45	0.510	0.09	trivial
Fatima	0.40	0.502	0.19	small

* Evaluation of the effect is based on Cohen' suggestion (1988), as follows: if d is less than 0.2, the effect should be considered trivial; if 0.2 ‹ d ‹ 0.5 the effect size may be considered 'small', if 0.5 ‹ d ‹ 0.8 the effect size can be considered 'medium', and if d is over 0.8 this represents a 'large' effect size.

According to our hypothesis, the largest gender gap should be found in the case of Arabic males vs. Arabic females. Comparing the difference in terms of instances of positive feedback shows that in most cases females received more positive feedback than males within the same nationality, except for in the case of the Dutch testers (Luuk got 10% more positive feedback than Julia). The highest gender gap was found in case of the Arabic testers: Fatima received 20% more positive responses than Hassan (Table 3).

Table 3. *Characteristics of treatment effect on gender by nationality: Cohen's d, according to nationality of testers (females vs. males) (N=20 in case of all profiles)*

	Difference in means on positive outcome (in percent)	**Standardized mean difference** Effect size (Cohen's d)	Evaluation of the effect*
Dutch: Julia-Luuk	-10%	0.202	small
Russian: Natasha-Vladimir	5%	0.100	trivial
Chinese: Chen-Li	15%	0.314	small
Arabic: Fatima-Hassan	20%	0.447	small
All female- all male testers	**7.5%**	**0.152**	**trivial**

* Evaluation of the effect is based on Cohen' suggestion (1988), as follows: if d is less than 0.2, the effect should be considered trivial; if 0.2 ‹ d ‹ 0.5 the effect size may be considered 'small', if 0.5 ‹ d ‹ 0.8 the effect size can be considered 'medium', and if d is over 0.8 this represents a 'large' effect size.

Passengers' chances of being offered a ride: multivariate analysis (Logistic regressions)

We built different logistic regression models to test our hypothesis. Our main aim was to measure odd ratios for the different testers, which is also a useful way to express effect size in discrimination studies. We tested our model with an *extended* set and with a *restricted* set of predictors. As most of the predictors in the extended model – including drivers' characteristics – were not significant, in the main text we show the models that include predictors related only to testers' and route characteristics (extended models can be found in the Appendices, Table A2).

The definition and further explanation of the dependent variables and the predictors is summarized below. The dependent variable in the regression: positive (value=1) vs. other outcome (negative or no response=0)

The model predictors were the following, with reference categories underlined:
• Testers' profile (in four categories): <u>Dutch</u>, Chinese, Russian, Arabic
• Routes' characteristics:
  – Route type: binary variable (<u>within</u>/outside Hungary)
  – Number of free seats (1 to 4)
• Drivers' characteristics: only used in the extended model
  – Age-groups (in categories: <u>18-29</u>, 30-45, 46-61 yrs.)
  – Gender (<u>male</u>/female)

- "Chattiness" score (1, 2 or 3: the higher the score, the more chatty the driver, as self-reported on their profile)
- Number of reviews (most of these are overwhelmingly positive so we decided to use this instead of the review score, which in many cases was lacking).

The most important results from the model are included in Table 4.

Table 4. *Odds ratios of obtaining positive feedback according to tester profile (Simple logistic regression models separately for male and female testers)*

		Positive outcome for all testers (total sample)		Positive outcome for male testers (split sample)		Positive outcome for female testers (split sample)	
		(1-offered a seat in the car, 0-refused or non-response)		(1- offered a seat in the car, 0-refused or non-response)		(1- offered a seat in the car, 0-refused or non-response)	
		N=160	Adjusted R Square =11%	N=80	Adjusted R	N=80	Adjusted R
		Chi square= 13.6*	Model sign= 0.018	Chi square= 18.03**	Model sign= 0.003	Chi square= 4.5	Model sign= 0.469
		Sign. (Wald test)	odds ratio (Exp (B))	Sign. (Wald test)	odds ratio (Exp (B))	Sign. (Wald test)	odds ratio (Exp (B))
Testers' profile	Ethnicity (Dutch)	.079		.040		.871	
	Chinese	.218	.562	.470	.590	.954	.963
	Russian	.415	.685	.161	.354	.784	1.198
	Arabic	.011	.286*	.004	.083**	.576	.692
Route	Route type (Within Hungary)						
	Outside Hungary	.007	2.602**	.007	6.395**	.402	1.486
	Free places	.742	.931	.279	1.494	.100	.615
	Constant	.965	1.028	.282	.312	.268	2.545

$^{*} p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Based on Chi square tests, the overall model was significant for the total sample and for the male subsample (at the level of 0.05). In contrast, the effects of the predictors included in our model were not significant in case of the female subsample. In these models, the predictors explained 11% and 27% (Adjusted R square) of the heterogeneity of the dependent variables, respectively. It should be mentioned, however, that the high proportion of explained variance in case of the male subsample might be the result of the low number of observations (N=80). From the predictors, only ethnicity and route type were significantly correlated with the outcome (both in the simple and in the extended models). Compared to the Dutch tester, the Arabic tester had an extremely low odds ratio in the male subsample (= 0.083) in terms of obtaining positive responses from drivers (in the total sample the odds ratio for the Arabic profile was also relatively low: 0.286). Interestingly, the only other significant predictor is type of route (both in the extended and the simple models): when the destination was outside of Hungary, the chance of being accepted for the ride was much higher compared to within Hungary, again, when other predictors in the model were controlled for.

In summary, we found the greatest differences within male testers' chances: the odds ratios for young Arabic male passengers being offered a ride were significantly less than those for the Dutch tester, which served as benchmark in our study.

CONCLUSION AND DISCUSSION

Our major hypothesis about homophily between ingroup and outgroup was partially proven by our small-scale experimental study; this assumption held when male and female outgroups were separately analysed.

As far as distrust towards male outgroups is concerned, the data showed that the highest rate of approval applied to the case of the Dutch and the lowest rate of approval the Arabic male tester. We assume that the Arab male profiles being significantly more rejection-prone (and receiving fewer responses) might be a consequence of heightened distrust towards Arab people, especially males. As far as distrust towards female outgroups is concerned, differences by nationality are not that great.

The hypothesis about the gender gap – which expects higher response and approval rates for women than for men due to a stronger tendency to trust women – was not proven statistically. Although we found a slight difference in the proportion of positive feedback rate by gender (38.7% for males vs.

46.3% for female) and we found different levels of gender gap (measured by the difference in means for positive outcomes) in the case of all three experimental tester pairs (the greatest gender gap was found in case of the Arabic testers), these differences were not significant.

We assume, however, that the lack of significance of the gender gap might be caused by a Type II error which can be tested by increasing the sample size in future tests using the same experimental design.

Finally, the multivariate analysis suggested clear signs of interaction between gender and perceived race: by all measures the Arabic male tester had the lowest rate of acceptance, the lowest odds ratio, and had to wait the longest time for answers, whereas the Dutch male tester had the highest chance of being offered a lift and waited the shortest time for the driver's answer. Moreover, in the case of the three experimental testers, males got less positive feedback. The overwhelming majority (94%) of drivers were male, rendering a general analysis of homophily between drivers and passengers in terms of gender invalid.

As far as validity issues are concerned, we believe that our research design had high internal validity as we were able to control all aspects of the driver-passenger interaction, using standardized profiles for all accounts and messages that varied only slightly. One limit to our study's internal validity is that we did not rely on computer-generated images to control for similarity, but instead searched for and chose images according to our subjective views of similarity across race and gender. We minimized this effect by showing the photos to several researchers before making a final selection. We purposefully only used the English language for communications, although we were aware that many drivers would self-select not to communicate, in order to compare the difference in racial/nationality groups.

The research has low external validity, and the findings can only be said to represent the ridesharing sector of the sharing economy. Different manifestations of the sharing economy rely on different interactions – some operate without financial incentives, and others require less trust than ridesharing, limiting the generalizability of our findings. Another limit to the external validity was that, in order to satisfy our research model, we did not interact with drivers who accepted automatic bookings. This likely prevented interaction with the most trusting subset of drivers on the site: those who accept passengers without viewing their profiles.

The final limit to the external validity of our findings is that we were only able to test what happened in the online sphere of the sharing economy rather than in real-life interactions. However, much of the sharing economy is organized and executed online.

REFERENCES

Airbnb (2017), "The economic impacts of home sharing in cities around the world". Last retrieved 8 August 2017 from: https://www.airbnb.com/ economic-impact

Arrow, Kenneth (1998), "What has economics to say about racial discrimination?", *The Journal of Economic Perspectives* Vol. 12, No 2, pp. 91-100. https://doi. org/10.1257/jep.12.2.91

Becker, Gary (1971), *The Economics of Discrimination*, Chicago, The University of Chicago Press

Belk, Russell (2014), "You are what you can access: sharing and collaborative consumption online", *Journal of Business Research* Vol. 67, No 8, pp. 1595-1600. https://doi.org/10.1016/j.jbusres.2013.10.001

Codagnone, Cristiano – Bertin Martens (2016), "Scoping the sharing economy: origins, definitions, impact and regulatory issues", *Institute for Prospective Technological Studies Digital Economy Working Paper 2016/01*, JRC100369. Last retrieved 8 August 2017 from: https://ec.europa.eu/jrc/sites/jrcsh/files/JRC100369.pdf

Cohen, Jacob (1988), *Statistical Power Analysis for the Behavioral Sciences*, Routledge

Dén-Nagy, Ildikó – Gábor Király (2014), "How to explain couchsurfing's success?", *Szociológiai Szemle* Vol. 24, No 4, pp. 32-53. Last retrieved 26 March 2018 from: http://www.szociologia.hu/dynamic/2014_4_den_nagy_ kiraly_32_53.pdf

Cui, Ruomeng – Jun Li – Dennis Zhang (2016), "Discrimination with Incomplete Information in the Sharing Economy: Evidence from Field Experiments on Airbnb". Last retrieved 26 March 2018 from: https://ssrn. com/abstract=2882982

Eagly, Alice – Maureen Crowley (1986), "Gender and helping behavior. A meta-analytic review of the social psychological literature", *Psychological Bulletin* Vol 100, No 3, pp. 283–308. http://dx.doi.org/10.1037/0033-2909.100.3.283

Edelman, Benjamin – Michael Luca (2014), "Digital discrimination: the case of Airbnb", *Harvard Business School,* Working paper 14-054. https://doi. org/10.2139/ssrn.2377353

Ert, Eyal – Aliza Fleischer – Nathan Magen (2016), "Trust and reputation in the sharing economy: the role of personal photos in Airbnb", *Tourism Management* Vol. 55, p. 62-73. https://doi.org/10.1016/j.tourman.2016.01.013

Ge, Yango – Christopher R. Knittel – Don MacKenzie – Stephen Zoepf (2016), "Race and Gender Discrimination in Transportation Network Companies", *The National Bureau of Economic Research*, Working Paper No. 22776. http:// dx.doi.org/10.3386/w22776

Gerber, Alan – Donald Green (2012), *Field Experiments: Design, Analysis, and Interpretation*, New York, WW Norton

Hamari, Juho – Mimmi Sjöklint – Antti Ukkonen (2016), "The sharing economy: why people participate in collaborative consumption", *Journal of the Association for Information Science and Technology* Vol. 67, No 9, pp. 2047-2059. https://doi.org/10.1002/asi.23552

Morton, Rebecca B. – Joshua A. Tucker (2014), "Experiments, Journals, and Ethics", *Journal of Experimental Political Science* Vol. 1, No 2, pp. 99–103. doi:10.1017/xps.2014.23

Koopman, Christopher – Matthew D. Mitchell – Adam D. Thierer (2015), "The sharing economy and consumer protection regulation: the case for policy change", *The Journal of Business, Entrepreneurship & the Law* Vol. 8, No 2, pp. 529-545. http://dx.doi.org/10.2139/ssrn.2535345

MacDonald, Andrew (2014), "UberDATA: Uber's economic impact on the city of Chicago", *UBER Blog*. Last retrieved 8 August 2017 from: https://www.uber.com/blog/chicago/uberdata-ubers-economic-impact-on-the-city-of-chicago/

Massey, Douglas S. – Garvey Lundy (2001), "Use of Black English and Racial Discrimination in Urban Housing Markets: New Methods and Findings", *Urban Affairs Review* Vol. 36, No 4, pp. 452–69. doi:10.1177/10780870122184957.

Mazzella, Frédéric – Arun Sundararajan (2016), "Entering the Trust Age." *Bla Bla Car/NYC Stern.*

Last retrieved 23 March 2018 from: https://www.blablacar.com/wp-content/uploads/2016/05/entering-the-trust-age.pdf

Pálosi, Éva – Endre Sik – Bori Simonovits (2007), "Diszkrimináció a plázában", *Szociológiai Szemle* Vol. 17, No 3–4, pp. 135–148.

Resnick, Paul - Richard Zeckhauser (2002), "Trust among strangers in internet transactions: empirical analysis of eBay' s reputation system", in Baye, Michael R., ed., *The Economics of the Internet and E-commerce (Advances in Applied Microeconomics, Vol. 11),* Emerald Group Publishing Limited, pp. 127 – 157.

Saucier, Donald A. – Carol T. Miller – Nicole Doucet (2005), "Differences in Helping Whites and Blacks: A Meta-Analysis", *Personality and Social Psychology Review* Vol. 9, No 1, pp. 2–16. https://doi.org/10.1207/s15327957pspr0901_1

Shaheen, Susan – Daniel Sperling – Conrad Wagner (1999), "A short history of carsharing in the 90's", *The Journal of World Transport Policy and Practice* Vol. 5, No 3, pp. 18-40.

Shoben, Elaine W. (1981), "Compound Discrimination: The Interaction of Race and Sex in Employment Discrimination", Scholarly Works. Paper 582. Last retrieved 26 March 2018 from: http://scholars.law.unlv.edu/facpub/582

Sik, Endre – Bori Simonovits (2008), "Egyenlő bánásmód és diszkrimináció", in Kolosi, Tamás & Tóth, István György, eds., *Társadalmi Riport 2008,* Budapest, TÁRKI, pp. 363–386.

Sztompka, Piotr (2003), Trust: A Sociological Theory, Cambridge, Cambridge University Press

APPENDIX: SUPPLEMENTARY MATERIAL FOR THE FIELD EXPERIMENT

Illustration of request and reply

Request:

- Hello! I would like to join you in travel from Budapest to ... *(Debrecen, Szeged, Bratislava, Graz).* I won't have any luggage with me. Can you say please, where is the meeting point in Budapest and do you still have one free place?
- Hi! I would like to travel with you to *(Debrecen, Szeged, Bratislava, Graz)* from Budapest. I have only a small backpack. Please let me know, where is the meeting place in Budapest and if you still have one free place?

Reply:

- In the case of rejection: Thank you. I will search for other option. Have a nice day.
- In the case of approval: Thank you for information, but unfortunately, I need to postpone my trip. Have a nice day.

Table 1. Attribute variables and experimental design in the pilot research (Hungary, 2017)

Attributes	Attribute level
Data collection	15th June - 30th July 2017 (weekdays and weekends)
Ridesharing platform	International ridesharing platform
Routes	8 internal routes within Hungary (Szeged, Debrecen, Sopron, Miskolc, Győr, Nyíregyháza and Pécs) 5 external routes (crossborder routes towards neighboring countries: (Bratislava, Vienna, Graz, Oradea, Novi Sad and Arad))
Nationality (experimental stimulus)	Dutch (benchmark)/Russian/Chinese/Arabic
Gender (experimental stimulus)	Men and women
Social status (control variable)	Homogenous (university students)

Table A1. *Mean and std. dev. for lag between passenger's request and driver's response (minutes)*

Passenger's name	Mean	Std. deviation	Min	Max
Vladimir	682,90	1260,55	13	3380
Luuk	292,79	407,72	1	1237
Li	586,55	1030,82	2	3457
Hassan	1531,42	2411,17	10	8085
Natasha	976,88	1857,36	1	7437
Julia	823,23	1293,07	6	4112
Chen	672,00	955,80	23	3410
Fatima	458,80	832,36	4	3168

Table A2. *Odds ratios of getting positive feedback according to tester profiles (Logistic regression models separately for male and female testers)*

		Positive outcome for male testers		Positive outcome for female testers	
		(1-offered a seat in the car, 0-refusal or non-response)		(1-offered a seat in the car, 0-refusal or non-response)	
		N=79	Adjusted R Square=27%	N=80	Adjusted R Square=9%
		Chi square=25.217	**model sign=0.014**	Chi square=8.51	model sign=0.744
		Sig. (Wald test)	odds ratio (Exp (B))	Sig. (Wald test)	odds ratio (Exp (B))
Driver's characteristics	Driver's nationality (Hungarian)	.990		.450	
	Driver's nationality: other than Hungarian	.889	.888	.306	.527
	Unknown	.919	.905	.656	1.841
	Driver Number of Reviews	.209	.972	.558	.981
	Drivers' Gender (male)				
	Drivers' Gender: female	.997	1.007	.309	.372
	Drivers' Age (18-29 yrs.)	.185		.526	
	Drivers' age (30-45 yrs.)	.067	.291	.321	.568
	Drivers age (46 yrs. or more)	.677	.720	.401	.555
	Chattiness score	.930	1.060	.848	1.107
Passenger's profile	Passengers' ethnicity (Dutch)	.041		.860	
	Chinese	.688	.725	.779	1.222
	Russian	.104	.254	.747	1.244
	Arabic	**.004**	**.066**	.656	.731
Route	Route type (Within Hungary)				
	Outside Hungary	**.010**	**10.785**	.331	1.693
	Free places	.085	2.141	.121	.595
	Constant	.438	.234	.419	3.885