I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11460

Help, Prejudice and Headscarves

Luis Artavia-Mora
Arjun Bedi
Matthias Rieger

APRIL 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11460

Help, Prejudice and Headscarves

Luis Artavia-Mora
ISS, Erasmus University of Rotterdam

Arjun Bedi
ISS, Erasmus University of Rotterdam and IZA

Matthias Rieger
ISS, Erasmus University of Rotterdam

APRIL 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Help, Prejudice and Headscarves*

This paper employs a natural field experiment in the Netherlands to test whether individuals intuitively help strangers with different group identities. We implement time manipulations in an everyday task to stimulate intuitive versus deliberate decision-making and thereafter examine helpfulness towards a female stranger with in-group (native) or out-group (Muslim) appearance. We find that time delay decreases helping rates. In contrast, regardless of time manipulation, out-group appearance does not influence helping rates. Overall, subjects are intuitively predisposed to help, independent of identity. We discuss our findings with respect to the literature on in-group favoritism and the cognitive origins of human cooperation.

JEL Classification: D03, D63, D64

Keywords: help, cooperation, in-group favoritism, Muslim, dual-process of cognition, natural field experiment, The Netherlands

Corresponding author:
Luis Artavia-Mora
International Institute of Social Studies
Erasmus University Rotterdam
Kortenaerkade 12
2518 AX, The Hague
The Netherlands
E-mail: artaviamora@iss.nl

* The ISS Research Ethics Committee approved the design and fieldwork of this study. We thank the Development Economics group at ISS for funding this research. We thank Zsófi Kelemen for data collection and fieldwork assistance and Robert Sparrow and Ian Levely for their comments. This paper is part of Luis Artavia-Mora's Ph.D thesis at ISS supervised by Dr. Matthias Rieger and Prof. Arjun Bedi.

1. Introduction

Humans prefer to interact and cooperate with people of their kind, a bias called in-group favoritism.[1] The origin of this bias has been attributed to the formation of cultural groups (Efferson et al. 2008) and linked to theories of group conflict, social identity and reputation-based models (Akerlof and Kranton, 2000; Tajfel and Turner, 1979; Tajfel, 1974; Becker, 1957). At the same time, a growing body of work has shown that prosocial behavior and cooperation lie at the core of socio-economic interactions in modern societies.[2] A reading of these two strands of work raises a natural question: does in-group favoritism crowd out prosocial behavior or does the desire to help dominate the effect of group identity? This paper relies on a natural field experiment to investigate this question. While there is an existing body of work which examines in-group favoritism and cooperative behavior in a lab setting, this paper provides novel evidence on the extent to which group identity influences pro-social decisions in anonymous one-time interactions in a natural field setting.[3] Such evidence is important for our understanding of in-group favoritism, human cooperation, and the evolution of modern societies.

To develop our experimental design we draw on the literature that focuses on the cognitive mechanisms governing human cooperation (Rand, 2016; Capraro and Cococcioni, 2016; Peysakhovich and Rand, 2016; Rand et al., 2014; Cone and Rand, 2014; Rand and Epstein, 2014; Rand and Nowak, 2013; Rand et al., 2012). This literature, which is based on the dual-process cognitive framework, investigates how two systems of human reasoning influence everyday decisions: (1) the intuitive system governs automatic, emotional and fast decisions, often linked to prior experience and beliefs and (2) the deliberative system comprises decisions that are more rational, reflective and slow (Evans and Stanovich, 2013; Kahneman, 2012; Frankish and Evans,

[1] See for instance reviews by Bertrand and Duflo (2017), Everett et al.(2015) and Anderson et al. (2006). Specifically, evidence of in-group favoritism has been found in labour markets (Weichselbaumer, 2016; Neumark, 2016; Baert et al., 2015; Pierné, 2013), housing markets (Bartoš et al., 2016; Auspurg et al., 2017; Ahmed and Hammarstedt, 2008), in public services (Giulietti et al., 2017;), other economic markets (Gautier et al., 2009; List, 2004; Riach and Rich, 2002), social dilemmas (Balliet et al., 2014; Nier et al., 2001), political choices (Rand et al., 2009; Beaton et al., 2008), and in evolutionary outcomes (Enke, 2017; Henrich, 2004).

[2] See for instance, Jordan et al. (2014), Capraro (2013), Gächter (2012), Bowles and Gintis (2011), Sigmund (2010), Nowak (2006), Fehr and Fischbacher (2004) and Axelrod (1984).

[3] Papers which have examined interactions between in-group favoritism and various social dilemmas in a laboratory setting include Chuah et al. (2016), Currani and Mengel (2016), Harris et al. (2015), Chuah et al. (2014), Chuah et al. (2013), Abbink and Harris (2012), Halevy et al. (2012), Chen and Li (2009) and Yamagishi and Mifune (2009; 2008).

2009; Loewenstein and O'Donoghue, 2004; Kahneman, 2003; Sloman, 1996). Typically, lab experiments implement time manipulations to investigate how the two systems influence decisions in the context of cooperation and discrimination. Time pressure is used to trigger intuitive choices and time delay to stimulate deliberate decision-making (Rand, 2016). There is extensive evidence favoring the argument that deliberation undermines intuitive predispositions to cooperate (Rand, 2016; Rand et al., 2012; 2013; 2014; Cone and Rand, 2014).[4]

Existing theories predict that preferences favoring the in-group develop during early childhood and become deeply engrained in later stages of life (Everett et al., 2017; Cikara and Van Bavel, 2014; Dunham et al., 2008). As a result, adults may be intuitively predisposed to favoring people of their own identity. However, recent lab experiments yield ambiguous evidence regarding this hypothesis. Consistent with theoretical expectations, De Dreu et al. (2015) and Ten Velden et al. (2017) report that intuition augments in-group favoritism. In contrast, Ma et al. (2015) find that (priming) intuition decreases in-group preference while reflection magnifies it. Rand et al. (2015) document that intuition increases cooperation across groups and in different social contexts, but has no effect on in-group favoritism. Everett et al. (2017) report a similar finding.

In this paper we contribute to the literature by investigating whether time delay influences the likelihood of helping strangers with different group identities using a natural field experiment in The Netherlands.[5] The paper is to the best of our knowledge the first to investigate the cognitive underpinnings of in-group favoritism in a natural task in the field. While we consider pure helping behavior with no further strategic component, our paper is located within the general framework of intuitive cooperation.

[4] The existing evidence does not support an unambiguous conclusion. One group of studies finds no clear evidence for the case of intuitive prosociality (Tinghög et al., 2013; Verkoeijen and Bouwmeester, 2014). Some studies show that this phenomenon is restricted to populations and environments with high levels of trust, or that cooperative predispositions are stronger among "inexperienced" participants (Capraro and Cococcioni, 2016; Rand et al., 2015; Rand and Kraft-Todd, 2014; Rand et al., 2013). Theoretically, Bear and Rand (2016) have reconciled these findings by showing an evolutionary equilibrium of intuitive prosociality using a game theory model of cognitive cooperation. Jagau and van Veelen (2017) have recently expanded this model by providing a general evolutionary framework of a dual-cognitive process of cooperation.

[5] We follow the set-up in Artavia-Mora et al. (2017) which extended the designs by Balafoutas et al. (2014) and Balafoutas and Nikiforakis (2012).

Our approach has two special features as compared to lab experiments. First, our experimental participants are diverse in terms of socio-demographic characteristics. Second, we randomly and very naturally implement both the time manipulation and the identity treatment in the context of everyday human interactions. Instead of using a computer lab and imposing abstract group identities on participants, the experiment is designed such that participants may provide help, at a small personal cost, to a stranger. Thus, our experimental design is less susceptible to participants' mistakes and confusion (List, 2011; 2007).

In our helping task we request an actress to drop a glove in a public park, and we record whether those passing by help by returning the glove (i.e., pick-up the glove or alert the actress). To test the impact of in-group favoritism on helping rates in this natural setting, the actress had two different appearances. For the in-group, native, control condition, she looked like a "stereotypical" Western European woman with white skin, blond hair and casual western clothes. For the out-group treatment, we asked the same actress to cover her hair and shoulders with a headscarf signaling that she was Muslim. Headscarves may proxy the unwillingness of Muslims to adapt to European customs. In Western countries such as the Netherlands, headscarves are often associated with intolerant, violent and anti-social behavior (Weichselbaumer, 2016; Pazhoohi and Burriss, 2016; Helbling, 2014; Ghumman and Ryan, 2013; Van der Noll, 2010; Byng, 2010; Adida et al., 2010; Ghumman and Jackson, 2010; Shadid and van Koningsveld, 2005).

To examine the effect of time delay on helping rates we re-use the design and location from Artavia-Mora et al. (2017). We exogenously manipulate the time available to make a decision in order to trigger intuitive versus deliberate responses. We asked our actress to drop the glove 4.5m or 13m away from subjects, giving them, on average, 3.5 seconds or 10 seconds, to come to a decision. The shorter distance promotes time pressure and is expected to prompt fast and intuitive responses, while the longer distance is our time delay treatment which is meant to elicit slow and deliberate decisions. In addition to recording helping rates, we also gathered information on basic socio-demographic characteristics from our participants using a post-experimental survey.

Our study tests three hypotheses. First, the headscarf treatment decreases helping rates compared to the native control condition (H1). Second, the time delay treatment decreases helping rates compared to time pressure (H2). Third, the interaction of time delay and the headscarf treatment magnifies out-group discrimination in helping rates (H3).

We find that the headscarf treatment has a negative but insignificant effect on helping rates and that participants are intuitively predisposed to help strangers regardless of group identity. The headscarf treatment reduces helping rates by -6.7%-points. However the effect is small relative to helping rates in the control condition (74.7%) and is insignificant (p-value=0.19, n=310). In contrast, time delay substantially decreases helping rates by 22.4%-points (p-value=0.00, n=310). Effects are statistically similar in the headscarf treatment and native condition. The coefficient associated with the interaction of time delay and headscarf is negative and insignificant (6.1%-points, p-value=0.54, n=310). In sum, the experiment yields strong evidence of intuitive help (Hypothesis 2) but no evidence that the headscarf treatment impacts helping rates (Hypothesis 1) or that time delay substantially magnifies defection in the headscarf treatment (Hypothesis 3).

The paper is organized as follows: Section 2 details the experimental design, the data collection and the balance of participant characteristics across treatments. Section 3 presents and discusses results, and Section 4 concludes.

2. Experimental Design

This section outlines the experimental design including task, location, treatments, as well as procedure and participant selection.

2.1 Experimental Task

Our aim is to investigate the combined impact of wearing a headscarf and time manipulations on pro-social behavior in a natural field setting. We assess pro-social behavior by measuring helping rates towards strangers in one-shot interactions. The experiment instructed an actress to drop (lose) a glove in a public park. We then record whether individuals who are passing by help or defect, and we interpret this choice as the decision to benefit a stranger at a personal cost (to the participant) or selfishness. The choice of a glove drop rather than another item is based on five

reasons (Artavia-Mora et al., 2017). First, it is a likely everyday event in a public space. Second, the usefulness of a single glove is limited. Third, a glove minimizes competing motivations to help or defect due to the monetary value of the item. For instance, losing a luxury item may incentivize theft, while losing an inexpensive one may be perceived as littering. Fourth, gloves can be dropped inaudibly, and thus, it is credible that the actress drops the glove without noticing. Fifth, a dropped glove is visible from the distances used in the experiment, an important condition for the time manipulation treatment explained below.

2.2 Location

We ran trials on the main walking path of Park Malieveld in The Hague (The Netherlands). The location was used in our recent study on helping rates and time delay (Artavia-Mora et al., 2017). This new experiment allows us to assess the impact of wearing a headscarf and to replicate the findings of the previous study, keeping locational effects and task fixed. As detailed in Artavia-Mora et al. (2017), the location has three attractive features. First, subjects cannot easily avoid or dodge the experimental task as the pedestrian path is lined by trees and straight. There is only one path crossing the park at this location (review Annex Photos A.1 to A.4). Subjects walking down the path benefit from good visibility and there is little distraction (Annex Photo A.2 and Photo A.3). Second, on the basis of prior observation we noted that individuals tend to cross the park by themselves and there is substantial distance between pedestrians. This feature allows us to observe private and anonymous behavior. Third, the park is centrally located, next to the main train station and within walking distance of many public institutions, office buildings, and shopping areas. This setting yields a diverse sample of participants for the experiment.

2.3 Treatments

The experiment features two treatments and their interaction:

1. *Group identities: The headscarf treatment*

In the primary treatment we arbitrarily assigned two outfits to the actress: (i) For the native or control condition, the actress had the appearance of a “stereotypical” Western European woman, with white skin, blond hair and attired in casual western clothes; (ii) For the headscarf treatment, the same actress covered her hair and shoulders (review Photos 1 and 2 below).

Photos 1 and 2: - Control (left panel) and headscarf treatment (right panel)

Note: The two group conditions compare the white-skinned, blond-haired actress (control) versus the same actress wearing a headscarf that covers her hair and shoulders (treatment).[6]

2. Time manipulations: The time delay treatment

To test intuitive versus deliberate responses to the headscarf treatment, this study manipulated the time needed to decide whether to help or not by varying the distance between the actress and the participant (Figure 1). We used a short and a long distance which may be converted into seconds based on the average human walking time of 1.3m/s (Mohler et al., 2007). The short distance (4.5 m) triggers fast and time-pressured decisions as subjects have about 3.5s to decide whether to help or defect (Photos 3 and 4). The long distance (13 m) elicits slow and time-delayed decisions as subjects have about 10s (Photos 5 to 6). The same distances were successfully used in our previous experiment and they were calibrated to ensure good visibility of the glove drop in our location.

2.4 Experimental Procedure

The experiment was implemented between May and July of 2017. The data collection took place on working days (between 10am and 5pm) except on days with bad weather conditions (i.e. rain, storm, extreme wind and low temperature). We used the same actress for all trials. The actress was unaware of the research hypotheses being tested. She had no formal training in acting except for a few classes in theater and drama performance.

The researcher (first author of the paper) placed himself far enough to avoid social and reciprocity pressures and recorded subjects' behavior (Photo 6). Each experiment started when

[6] The actress has provided written permission to use her photos.

the researcher had chosen a participant based on two criteria. First, the participant had to be walking alone and far from others to ensure “private” decision making. Second, the participant was not visibly distracted (i.e. on a phone) or in a rush (i.e. jogging). Photos 7 and 8 show a typical participant on location. In rare cases, participants were excluded as they were known to the research team.

Figure 1 illustrates the experiment. The participant is shown in black and the actress in grey (in this case featuring the headscarf treatment). Each trial began with the actress holding the gloves and sitting on a bench. Bicycles marked points A, B or C (Photos 3 to 6). The actress wore the headscarf (or the control, native outfit) for arbitrary blocks of subjects. She then left the bench to cross the path towards Point A when a selected participant was approaching point B (or C, depending on the distance -time manipulation- treatment). When the participant arrived at either point B (or C) the actress dropped one glove, pretending not to notice the loss. She then stayed in front of her bicycle at Point A facing the open space while pretending to look for keys (Photos 3 to 6). Most importantly, the actress was instructed to react only at point A, always waiting for the participant to first help or defect at point A. Thus, the actress ignored all voice alerts far from Point A, guaranteeing that all subjects had the same time to help her. This strategy guaranteed an approximately fixed decision time across treatment groups. The behavior displayed by our participants include: “picking glove and giving back”, “signals and voice alert” and “only voice alert” (recorded as helping “1”) versus defection (recorded as selfishness “0”). Examples of helping behavior and defection are shown in Photos 9 to 11.

The actress also conducted a short, post-experimental survey after each participant had made the choice at Point A (Photo 12). The survey covered social and demographic characteristics of each participant including age, gender, time lived in The Netherlands, risk-taking behavior, body height, and religion (the survey instrument is in the Annex). This information is used to provide a sense of the nature of the subject pool, check treatment balance, and investigate treatment heterogeneities.

2.5 Participants Characteristics and Treatment Balance

Based on ex-ante power calculations we ran a total of 372 trials.[7] Using the post-experimental survey, Table 1 shows that the characteristics of participants are statistically balanced across all treatment conditions. Overall, the average participant was 45 years old and 41% were female. About 82% of the participants had always resided in The Netherlands (natives). We also enquired if subjects had seen the glove drop - 83% of the subjects acknowledged seeing it with no statistical differences across treatments. Our main analysis excludes participants who indicated that they had not seen the glove drop. As will be shown later in the text, the results are not sensitive to inclusion or exclusion of these cases. Finally, non-response to the survey was statistically unrelated to the treatments.[8] Table 2 displays the total sample sizes by group identity and time manipulation treatments.

3. Results

In this section we first present the baseline results and then investigate the robustness to including covariates. The section concludes by examining heterogeneities relating to the religion of the subject. For the analysis, time pressure and the native looking female are taken as the baseline conditions (control), while time delay and the headscarf are taken as treatment conditions.

3.1 Main Results

In total, 71% of the subjects helped the actress.[9] In comparison, Artavia-Mora et al. (2017) found overall helping rates of 66% in the same setting for a different actress,[10] while Balafoutas et al. (2014) found that 39.7% assisted actors of both sexes who dropped a book in a German train station.[11] Turning to our first hypothesis, we do not find statistically significant evidence to

[7] The main aim was to detect the impact of wearing a headscarf on helping rates. Aiming for a minimum detectable impact of 15%-points (p1=60%, p2=45%) for a two-sided test of proportions, and setting significance level to 5% and power to 80%, yielded a required sample size of 372. The power calculations were estimated using the STATA command *sampsi*.

[8] The survey response rate was 73%, and non-response is statistically unrelated to the experimental treatments (see p-values in Table 1).

[9] Table B1 in the Annex presents the percentages of type of helping behavior by treatments. There are three categories of helping behavior. These include, picking glove and giving back, signals and voice alert, and only voice alert.

[10] Differences in helping rates across the two samples are statistically insignificant (Δ5.4%-points, p-value=0.26, n=439).

[11] The fact that the actor was aware of dropping the book could have decreased the underlying motivations of the participants to help.

support the idea that wearing a headscarf decreases helping rates. Panel A of Figure 2 documents that a majority of participants helped the actress regardless of her outfit, that is, 74.68% and 67.95% for the native and headscarf look, respectively. The impact of wearing a headscarf amounts to a small and statistically insignificant reduction of -6.73%-points (p-value=0.19, n=310).[12] This estimate represents a small effect size of 0.15 according to Cohen's H for proportions.

Panel B of Figure 2 breaks down helping rates across actress outfits and time treatments. As predicted by hypothesis H2, time delay reduces helping rates by 19.48%-points (p-value=0.01, n=154) in the native and by 25.58%-points (p-value=0.00, n=156) in the headscarf treatment, respectively.[13] The relative reduction is large (in absolute terms) in the latter condition, however the interaction term between time delay and the headscarf treatment is statistically insignificant (-6.1%-points, p-value=0.54, n=310). Similarly, the impact of the headscarf treatment is negative but statistically insignificant in both the time delay (-10.27%-points, p-value=0.20, n=152) and time pressure conditions (-4.17%-points, p-value=0.50, n=158).

In sum, while the impact of the headscarf treatment on helping rates is consistently negative, suggesting prejudice against Muslims, the effect is small and statistically insignificant. To put the effect in perspective, in the baseline native condition 115 subjects helped the actress compared to 106 in the headscarf condition. There is also no strong impact associated with the interaction between the headscarf treatment and time delay.

3.2 Pooled Regression Model

Table 3 presents linear regression estimates of the time delay and headscarf treatment effects on the likelihood of helping the actress. As shown in Column 1, time delay leads to a substantial and statistically significant 22.4%-point reduction in the helping rate. The effect associated with the headscarf treatment is -6.7%-points, but insignificant and modest in size relative to the time

[12] With a significance level of 0.05 and power of 0.80, we would need a sample size of 1,478 observations for a minimum detectable impact of 6.73%-points corresponding to the impact shown in Figure 2, Panel A. This would involve an unfeasibly large experiment given design and location. The calculation was performed in STATA using the *sampsi* command.

[13] The average impact of time delay is -22.41%-points (p-value=0.00, n=310). We also asked respondents if it was difficult to decide whether to help or not. Respondents in the time delay treatment experienced significantly greater difficulty and were 7.1%-points more likely to respond that they found it "Quite" or "A lot" difficult (p-value=0.01; n=220).

delay impact and overall helping rates (Column 2). The coefficients remain stable when both treatments are jointly introduced (Column 3). The impact of time delay is significantly larger in (absolute) terms than the one associated with the headscarf (p-value=0.03). Finally, Column 4 provides estimates of the interaction between the two treatments. The coefficient on the interaction term is negative, indicating that time delay does magnify prejudice against Muslims. However, the effect is small and statistically insignificant (-6.1%-points, p-value=0.54). Under time pressure and when facing a female with a headscarf, subjects reduce helping rates by 4.2%-points, but when experiencing both a time delay *and* a female with a headscarf, participants reduce helping rates by 10.3%-points. However, these differences are imprecisely estimated as both the main effect of the headscarf treatment and the corresponding interaction terms are insignificant at conventional levels. Overall, the evidence does not support hypothesis H3.[14]

3.3 Robustness

Table 4 presents robustness checks by including participant covariates obtained from the post-experimental survey. An issue which we explore below is that "missing" data due to survey non-response may itself be correlated with helpfulness and the headscarf treatment. For instance, subjects may decline to be interviewed by the actress if they did not help her previously or if she is wearing a headscarf.

Column 1 repeats the baseline estimates from Table 3. Column 2 provides estimates restricted to the sample for which covariates are available (subjects who responded to the survey). The time delay estimate declines in magnitude but remains negative and statistically significant (-17.6%-points). The coefficient on the headscarf treatment is now positive but remains statistically insignificant. In column 3 we add subject covariates. The main estimates are not sensitive to their inclusion. With one exception all the covariates are statistically insignificant. The risk-taking measure is positively correlated with the likelihood of helping. An increase of 1 point in risk-taking is associated with a 3.8%-point (p-value=0.05) increase in the likelihood of helping the actress across both dilemmas. In column 4 we add the interaction term of the time delay and headscarf treatments. The treatment effect associated with time delay remains negative and sizeable but becomes imprecisely estimated while the effect of the headscarf treatment remains

[14] Table B2 presents results based on the full sample, including people that reported that they had not seen the glove drop. Results are statistically similar.

insignificant. Similar to the estimates based on the full sample the interaction term is negative indicating that time delay reduces the probability of helping a female with headscarf but the effect is small and statistically insignificant.

The volatility of the (insignificant) headscarf treatment effect may be due to sample selection, that is, systematic non-response to the post-experimental survey. We investigate this issue further in Annex Table B3 where we include a dummy variable for non-response in the regressions. The coefficient on the non-response dummy is negative and significant and indicates that those who did not respond to the survey were also less likely to help the actress. In other words, the restricted sample for which we have covariates consists of subjects who are more likely to help the actress. In columns 2 and 3, we add covariates and fill in missing values with sample means. The estimates associated with both treatments are now stable suggesting that the modest sensitivity in the estimates of Table 4 may be attributed to a systematic pattern in survey non-response.

We also hypothesized that subjects' responses to the headscarf treatment may depend on their own religious faith. For instance, a Muslim subject could have different cognitive reactions when helping a female with a headscarf in comparison to the reactions of a non-Muslim subject. One limitation is that 62% of subjects reported that they did not adhere to a religious faith and only 5% were Muslim. Keeping these figures in mind, Table 5 shows that a subject's religious faith had no significant effects on the likelihood of helping. Although participants' faith is statistically uncorrelated with overall helping behavior (column 1), Muslim participants were more likely to help the actress when she was wearing a headscarf (column 2). The effect is precisely estimated.[15]

4. Conclusion

This paper contributes to the literatures on (i) in-group favoritism and (ii) the cognitive origins of human cooperation. It provides evidence on the effect of group identity on helping behavior in a

[15] Main findings are robust to dropping Muslim participants from the sample. In the case of a sample which excluded Muslims, the coefficient associated with time delay remains precisely estimated and stable (-17.17%-points, p-value=0.00, n=214). The headscarf coefficient remains negative and insignificant (-4.36%-points, p-value=0.57, n=214).

natural field setting and explores its underlying cognitive mechanism. The study employed an actress with in-group or out-group markers and analyzed whether strangers return a lost item. For the in-group condition the actress (with white skin and blond hair) wore casual clothes, and in the treatment, she wore a headscarf to signal out-group membership in the European context (Weichselbaumer, 2016). To evaluate the cognitive drivers of in-group favoritism we manipulated the decision time available to participants. This design was used to test three hypotheses: (H1) the headscarf treatment reduces helping rates, (H2) that time delay reduces helping rates, and (H3) that time delay magnifies the effect of wearing the headscarf.

The experimental framework of our study allows us to examine prosocial choice and cognitive mechanisms in an everyday, private and anonymous one-shot interaction. This aspect differentiates our work from standard laboratory-based approaches involving altruism, cooperation and other social dilemmas (i.e. Everett et al., 2017; Chuah et al., 2016; Currani and Mengel, 2016; Rand et al., 2015; Harris et al., 2015; Chuah et al., 2014; Abbink and Harris, 2012). We naturally marked out-group identity using a headscarf treatment rather than artificial assignments based on team colors/names or announced out-group markers (i.e. religious beliefs, ethnicity or political affiliation). At the same time, our study expands the large literature on discriminatory behavior in labor, housing and other market-based interactions (i.e. Weichselbaumer, 2016; Neumark, 2016; Bartoš et al., 2017; Auspurg et al., 2017; Ahmed and Hammarstedt, 2008; List, 2004).

Our findings are in line with previous evidence showing that intuition increases pro-social choices. Also consistent with some previous evidence; the interaction between time manipulation and group membership is statistically insignificant (Everett et al., 2017; Rand et al., 2015).[16] Three other studies have found statistically significant interaction effects between cognition and group identity but they differ as they implement treatments based on oxytocin, cognitive taxation or priming, as well as combinations of these, instead of time manipulations (Ten Velden et al., 2017; De Dreu et al., 2015; Ma et al., 2015). In our study, intuitive helpfulness clearly dominates the main and interactive effects of group identity (in favor of H2, rejecting H1 and H3).

[16] In our research, the out-group treatment uses a headscarf. Instead, Everett et al. (2017) employ team assignment (Eagles or Rattlers) based on five personality questions, and Rand et al. (2015) use political support for Barack Obama or Mitt Romney in the 2012 US presidential election.

Our results could differ from previous evidence showing in-group favoritism in market and laboratory behavior, as we analyze a relatively low cost and easy everyday helping choice. It is possible that more costly decisions with an intertemporal or strategic dimension are also more discriminatory in nature (think of hiring personnel, renting an apartment or allocating money between people). Substantially increasing time delay could also change the results. Another point worth mentioning is that contexts and social groups with high levels of trust and pro-sociality can mediate the mechanisms driving human cooperation (Capraro and Cococcioni, 2016; Rand and Kraft-Todd, 2014; Rand et al., 2013). The Netherlands is internationally recognized for high levels of trust, honesty, cooperation and tolerance, especially towards immigrants and ethnic minorities. The city of The Hague is one of the most diverse and international cities in the country, home to embassies and international organizations such as the International Court of Justice. Thus, it is possible that our experiment will yield different results in other locations. Relatedly, it would be interesting to investigate how the degree of familiarity between group identities and/or spatial structures impact results (Mosleh and Rand, 2018).

The results reported in the paper inform work on in-group favoritism and dual-processes of cognition. First, future studies could investigate the stability of the findings as a function of location, experimental procedure, and actress traits. Second, future research could aim to understand how characteristics of the headscarf or using other group markers may lead to different effects. For instance, by examining the effect of more conservative out-group clothing such as a face veil or a full Muslim outfit (Pazhoohi and Burriss, 2016). Using more conservative clothing could plausibly lead to a larger and more pronounced impact of in-group favoritism on helping rates. Third, increasing the personal cost of the task or varying time delay could also shape decision-making. Fourth, laboratory studies should continue investigating the underlying cognitive mechanisms that drive intergroup interactions and pro-social behavior (Everett et al., 2017; Peysakhovich and Rand, 2016). This research will improve our understanding of when, why and how intuitive and deliberative decision-making influence social and economic interactions in modern societies.

5. References

Abbink, K. and D. Harris. 2012. *In-group favouritism and out-group discrimination in naturally occurring groups*. Working paper 616, University of Oxford.

Adida, C.L., D.D. Laitin, and M.A. Valfort. 2010. Identifying barriers to Muslim integration in France. *Proceedings of the National Academy of Sciences*, *107*(52), pp.22384-22390.

Ahmed, A.M. and M. Hammarstedt. 2008. Discrimination in the rental housing market: A field experiment on the Internet. *Journal of Urban Economics*, *64*(2), pp.362-372.

Akerlof, G.A. and R.E. Kranton. 2000. Economics and identity. *The Quarterly Journal of Economics*, *115*(3), pp.715-753.

Anderson, L., R. Fryer, and C. Holt. 2006. Discrimination: experimental evidence from psychology and economics. *Handbook on the Economics of Discrimination*, pp.97-118.

Artavia-Mora, L., A.S. Bedi, and M. Rieger. 2017. Intuitive help and punishment in the field. *European Economic Review*, *92*, pp.133-145.

Auspurg, K., T. Hinz, and L. Schmid. 2017. Contexts and Conditions of Ethnic Discrimination: Evidence from a Field Experiment in a German Housing Market. *Journal of Housing Economics*, *35*, pp.26-36.

Axelrod, R. 1984. The Evolution of Cooperation. Basic Books, New York.

Baert, S., B. Cockx, N. Gheyle, and C. Vandamme. 2015. Is there less discrimination in occupations where recruitment is difficult? *Industrial and Labour Relations Review*, *68*(3), pp.467-500.

Balafoutas, L. and N. Nikiforakis. 2012. Norm enforcement in the city: a natural field experiment. *European Economic Review 56*(8), pp.1773–1785.

Balafoutas, L., N. Nikiforakis, and B. Rockenbach. 2014. Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences*, *111*(45), pp.15924–15927.

Balliet, D., J. Wu, and C.K. De Dreu. 2014. In-group favoritism in cooperation: A meta-analysis. *Psychological Bulletin*, *140*(6), pp.1556.

Bartoš, V., M. Bauer, J. Chytilová, and F. Matějka. 2016. Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition. *American Economic Review*, *106(6)*, 1437-1475.

Bertrand, M. and E. Duflo. 2017. Field Experiments on Discrimination. *Handbook of Economic Field Experiments.* North-Holland, (1), pp. 309-393.

Beaton, A.M., J.F. Dovidio, and N. Léger. 2008. All in this together? Group representations and policy support. *Journal of Experimental Social Psychology*, *44*(3), pp.808-817.

Bear, A. and D.G. Rand. 2016. Intuition, deliberation, and the evolution of cooperation. *Proceedings of the National Academy of Sciences, 113*(4), pp.936–941.

Becker, G. 1957. The Economics of Discrimination. University of Chicago Press, Chicago.

Bowles, S. and H. Gintis. 2011. A cooperative species: Human reciprocity and its evolution. Princeton University Press, Princeton.

Byng, M.D. 2010. Symbolically Muslim: media, hijab, and the West. *Critical Sociology*, *36*(1), pp.109-129.

Capraro, V. 2013. A model of human cooperation in social dilemmas. *PLoS One, 8*(8), p.p.e72427.

Capraro, V. and G. Cococcioni. 2016. Rethinking spontaneous giving: extreme time pressure and ego-depletion favor self-regarding reactions. *Scientific Reports,* 6, pp.27219.

Chen, Y. and S.X. Li. 2009. Group identity and social preferences. *American Economic Review*, *99*(1), pp.431-457.

Chuah, S.H., R. Fahoum, and R. Hoffmann. 2013. Fractionalization and trust in India: a field-experiment. *Economic Letters*, *119*(2), pp.191–194.

Chuah, S.H., R. Hoffmann, B. Ramasamy, and J.H. Tan. 2014. Religion, ethnicity and cooperation: An experimental study. *Journal of Economic Psychology*, *45*, pp.33-43.

Chuah, S.H., S. Gächter, R. Hoffmann, and J.H. Tan. 2016. Religion, discrimination and trust across three cultures. *European Economic Review*, *90*, pp.280-301.

Cikara, M. and J.J. Van Bavel. 2014. The neuroscience of intergroup relations: An integrative review. *Perspectives on Psychological Science*, *9*(3), pp.245–274.

Cone, J. and D.G. Rand. 2014. Time pressure increases cooperation in competitively framed social dilemmas. *PLoS One*, *9*(12), pp.e115756.

Currarini, S. and F. Mengel. 2016. Identity, homophily and in-group bias. *European Economic Review*, 90, pp.40-55.

De Dreu, C.K., D.B. Dussel, and F.S.T. Velden. 2015. In intergroup conflict, self-sacrifice is stronger among pro-social individuals, and parochial altruism emerges especially among cognitively taxed individuals. *Frontiers in Psychology*, *6*, pp.572.

Dunham, Y., A.S. Baron, and M.R. Banaji. 2008. The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, *12*(7), pp.248–253.

Efferson, C., R. Lalive, and E. Fehr. 2008. The coevolution of cultural groups and ingroup favoritism. *Science*, *321*(5897), pp.1844-1849.

Enke, B. 2017. *Kinship Systems, Cooperation and the Evolution of Culture*. Working paper w23499, National Bureau of Economic Research.

Evans, J.S.B. and K.E. Stanovich. 2013. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), pp.223-241.

Everett, J., N.S. Faber, and M. Crockett. 2015. Preferences and beliefs in ingroup favoritism. *Frontiers in Behavioral Neuroscience*, 9, 15.

Everett, J., Z. Ingbretsen, F. Cushman, and M. Cikara. 2017. Deliberation erodes cooperative behaviour–even towards competitive outgroups, even when using a control condition, and even when controlling for sample bias. *Journal of Experimental Social Psychology*, 73, pp.76-81

Fehr, E. and U. Fischbacher. 2004. Social norms and human cooperation. *Trends Cognitive Sciences*, 8, pp.185–190.

Frankish, K. and J.S.B.T. Evans. 2009. The Duality of Mind: an Historical Perspective. Oxford University Press, Oxford, United Kingdom.

Gächter, S. 2012. A cooperative instinct. *Nature, 489*(7416), pp.374.

Gautier, P.A., A. Siegmann, and A. Van Vuuren. 2009. Terrorism and attitudes towards minorities: The effect of the Theo van Gogh murder on house prices in Amsterdam. *Journal of Urban Economics*, *65*(2), pp.113-126.

Ghumman, S. and A.M. Ryan. 2013. Not welcome here: Discrimination towards women who wear the Muslim headscarf. *Human Relations*, *66*(5), pp.671-698.

Ghumman, S. and L. Jackson. 2010. The downside of religious attire: The Muslim headscarf and expectations of obtaining employment. *Journal of Organizational Behavior*, *31*(1), pp.4-23.

Giulietti, C., M. Tonin, and M. Vlassopoulos. 2017. Racial Discrimination in Local Public Services: A Field Experiment in the United States. *Journal of the European Economic Association*, forthcoming.

Halevy, N., O. Weisel, and G. Bornstein. 2012. "In-Group Love" and "Out-Group Hate" in Repeated Interaction Between Groups. *Journal of Behavioral Decision Making*, *25*(2), pp.188-195.

Harris, D., B. Herrmann, A. Kontoleon, and J. Newton. 2015. Is it a norm to favour your own group? *Experimental Economics*, *18*(3), pp.491-521.

Helbling, M. 2014. Opposing Muslims and the Muslim headscarf in Western Europe. *European Sociological Review*, *30*(2), pp.242-257.

Henrich, J. 2004. Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior and Organization*, *53*(1), 3–35.

Jagau, S. and M. van Veelen. 2017. A general evolutionary framework for the role of intuition and deliberation in cooperation. *Nature Human Behaviour*, *1*(8), pp.0152.

Jordan, J.J., A. Peysakhovich, and D.G. Rand. 2014. Why we cooperate. The Moral Brain: Multidisciplinary Perspectives. MIT Press, Cambridge.

Kahneman, D. 2003. A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, *58*(9), pp.697.

Kahneman, D. 2012. Thinking, Fast and Slow. Penguin, London, United Kingdom.

List, J.A. 2004. The nature and extent of discrimination in the marketplace: Evidence from the field. *The Quarterly Journal of Economics*, *119*(1), pp.49-89.

List, J.A., 2007. Field experiments: a bridge between lab and naturally occurring data. *The BE Journal of Economic Analysis and Policy*, *6*(2).

List, J.A. 2011. Why economists should conduct field experiments and 14 tips for pulling one off. *Journal of Economic Perspectives*, *25*(3), pp.3–15.

Loewenstein, G. and T. O'Donoghue. 2004. *Animal Spirits: Affective and Deliberative Processes in Economic Behavior*. Working paper 04-14, Cornell University, Center for Analytic Economics.

Ma, Y., Y. Liu, D.G. Rand, T.F. Heatherton, and S. Han. 2015. Opposing oxytocin effects on intergroup cooperative behavior in intuitive and reflective minds. *Neuropsychopharmacology*, *40*(10), pp.2379–2387.

Mohler, B.J., W.B. Thompson, S.H. Creem-Regehr, H.L. Pick, and W.H. Jr. Warren. 2007. Visual flow influences gait transition speed and preferred walking speed. *Experimental Brain Research*, *181*(2), pp.221–228.

Mosleh, M. and D.G. Rand. 2018. *Thinking Fast and Slow on Networks: Co-evolution of Cognition and Cooperation in Structured Populations*. SSRN working paper.

Neumark, D. 2016. *Experimental research on labor market discrimination*. Working paper w22022, National Bureau of Economic Research.

Nier, J.A., S.L. Gaertner, J.F. Dovidio, B.S. Banker, C.M. Ward, and M.C. Rust. 2001. Changing interracial evaluations and behavior: The effects of a common group identity. *Group Processes and Intergroup Relations*, *4*(4), pp.299-316.

Nowak, M.A. 2006. Five rules for the evolution of cooperation. *Science*, *314*(5805), pp.1560–1563.

Pazhoohi, F. and P.R. Burriss. 2016. Hijab and "Hitchhiking": A field study. *Evolutionary Psychological Science*, 2(1), pp.32-37.

Peysakhovich, A. and D.G. Rand. 2016. Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science*, *62*(3), pp.631–647.

Pierné, G. 2013. Hiring discrimination based on national origin and religious closeness: results from a field experiment in the Paris area. *IZA Journal of Labor Economics*, *2*(1), pp.4.

Rand, D.G., T. Pfeiffer, A. Dreber, R.W. Sheketoff, N.C. Werner felt, and Y. Benkler. 2009. Dynamic remodeling of in-group bias during the 2008 presidential election. *Proceeding of the National Academy of Science*, 106, pp.6187–6191.

Rand, D. G., G.E. Newman, and O.M. Wurzbacher. 2015. Social context and the dynamics of cooperative choice. *Journal of Behavioral Decision Making*, *28*(2), pp.159–166.

Rand, D.G., A. Peysakhovich, G.T. Kraft-Todd, G.E. Newman, O. Wurzbacher, M.A. Nowak, and J.D. Greene. 2014. Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, pp.3677.

Rand, D.G. and T.G. Kraft-Todd. 2014. Reflection does not undermine self-interested prosociality. *Frontiers in Behavioral Neuroscience*, 8, pp.300.

Rand, D.G., J.D. Greene, and M.A. Nowak. 2012. Spontaneous giving and calculated greed. *Nature*, *489*(7416), pp.427–430.

Rand, D.G., J.D. Greene, and M.A. Nowak. 2013. Rand et al. reply. *Nature*, *498*(7452), pp.E3.

Rand, D.G. and M.A. Nowak. 2013. Human cooperation. *Trends Cognitive Sciences*, *17*(8), pp.413–425.

Rand, D.G. and Z.G. Epstein. 2014. Risking your life without a second thought: Intuitive decision-making and extreme altruism. *PloS One*, 9(10), pp.e109687.

Rand, D.G. 2016. Cooperation, fast and slow: meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Sci*ences, *27*(9), pp.1192-1206.

Riach, P. and J. Rich. 2002. Field experiments of discrimination in the market place. *The Economic Journal*, 112, pp.480-518.

Shadid, W. and P.S. van Koningsveld. 2005. Muslim dress in Europe: debates on the headscarf. *Journal of Islamic Studies*, *16*(1), pp.35-61.

Sigmund, K. 2010. The Calculus of Selfishness. Princeton University Press.

Sloman, S.A. 1996. The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), pp.3.

Tajfel, H. and J.C. Turner. 1979. An integrative theory of intergroup conflict. *The Social Psychology of Intergroup Relations*, *33*(47), pp.74.

Tajfel, H. 1974. Social identity and intergroup behaviour. *Information (International Social Science Council)*, *13*(2), pp.65-93.

Ten Velden, F.S., K. Daughters, and C.K. De Dreu. 2017. Oxytocin promotes intuitive rather than deliberated cooperation with the in-group. *Hormones and Behavior*, 92, pp.164-171.

Tinghög, G., D. Andersson, C. Bonn, H. Böttiger, C. Josephson, G. Lundgren, D. Västjäll, M. Kirchler, M. Johannesson. 2013. Intuition and cooperation reconsidered. *Nature 498*(7452), pp.E1–E2.

Van der Noll, J. 2010. Public support for a ban on headscarves: A cross-national perspective. *International Journal of Conflict and Violence*, *4*(2), pp.192.
Verkoeijen, P. and S. Bouwmeester. 2014. Does intuition cause cooperation? *PLoS One*, *9*(5), pp.e96654.

Yamagishi, T. and N. Mifune. 2008. Does shared group membership promote altruism? Fear, greed, and reputation. *Rationality and Society*, *20*(1), pp.5-30.

Yamagishi, T. and N. Mifune. 2009. Social exchange and solidarity: in-group love or out-group hate? *Evolution and Human Behavior*, 30, pp.229–237.

Weichselbaumer, D. 2016. *Discrimination Against Female Migrants Wearing Headscarves*. Working Paper 10217, Social Science Research Network.

6. Figures

Figure 1: Diagram of time manipulations

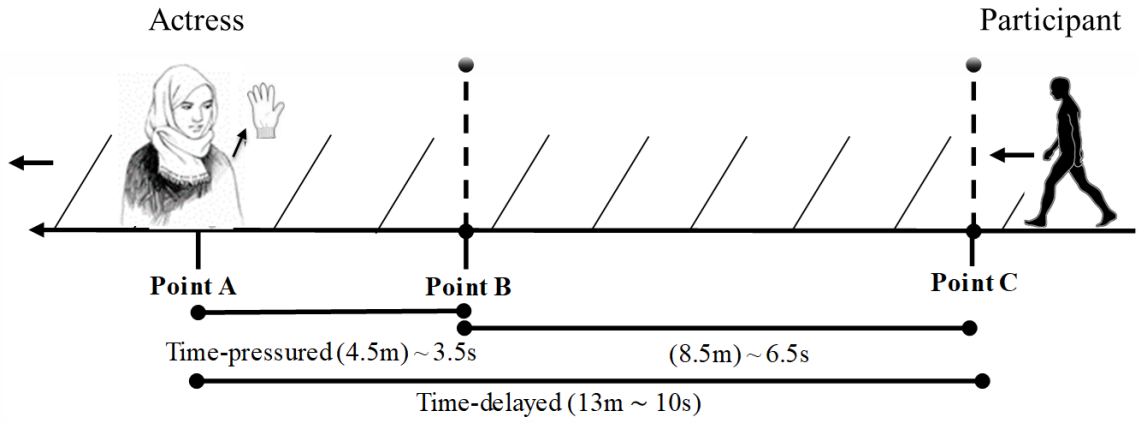
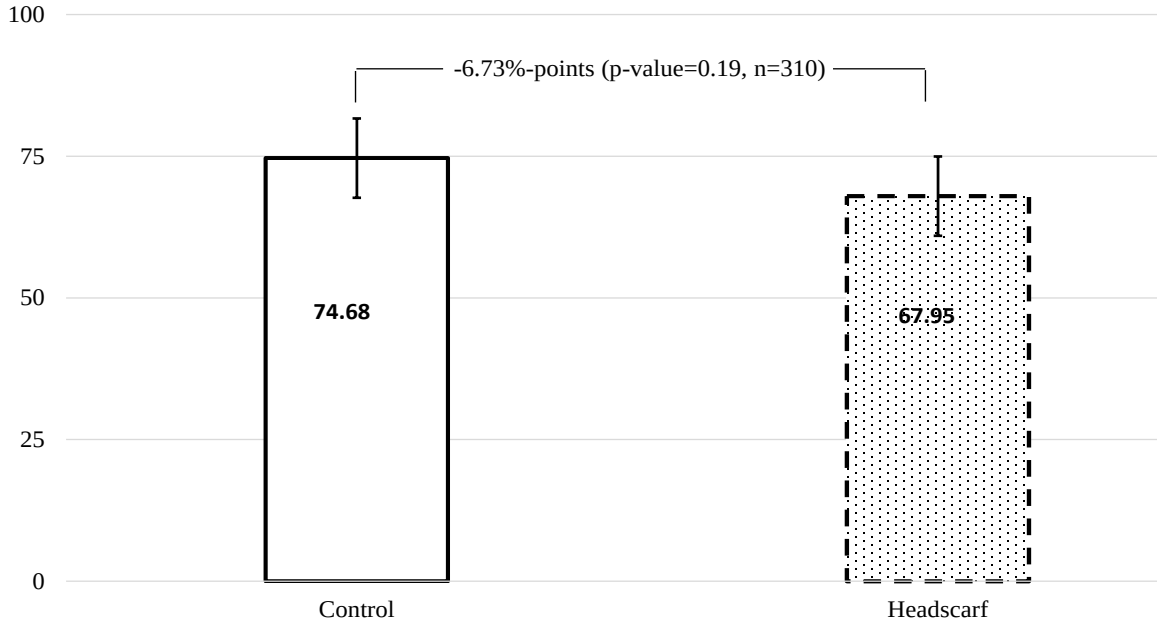
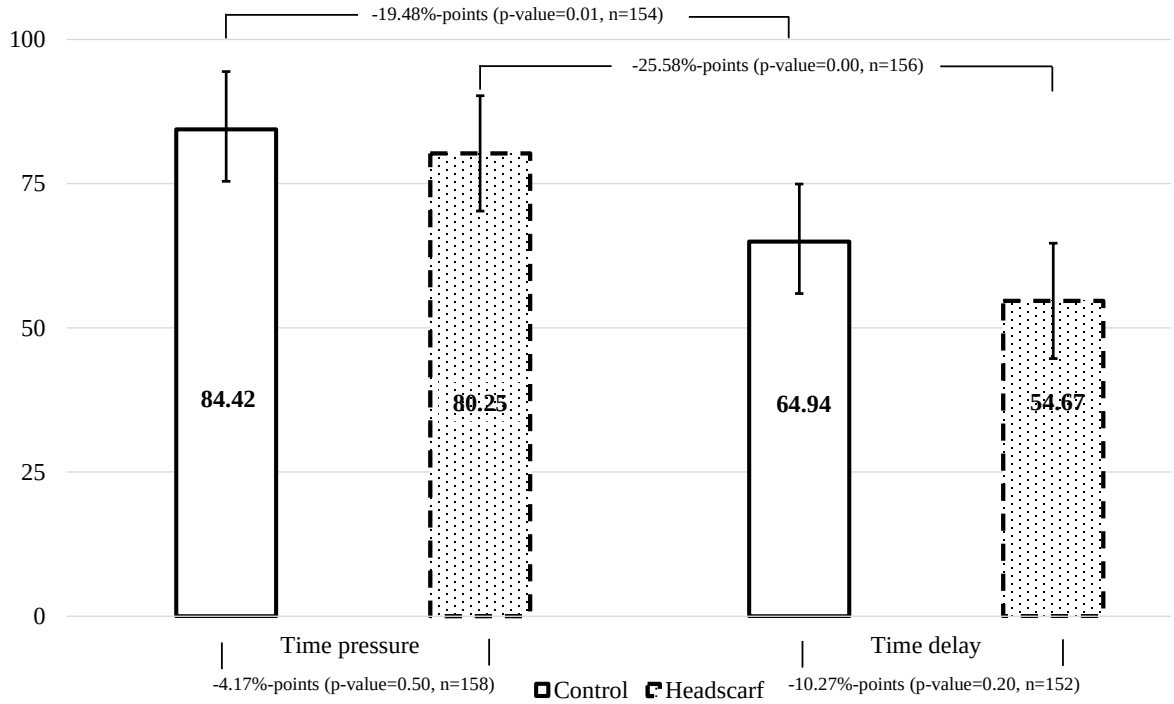


Figure 2: Average helping rates by female group identity and time manipulation (in %)

Panel A: Rates of helping by female group identity



Panel B. Rates of helping by time manipulation and female group identity



Note: P-values stem from robust two-sided t-tests.

7. **Tables**

Table 1: Subjects' characteristics and treatment balance (full sample)

						Treatments	
						Time	Headscarf
	N	Mean	SD	Min	Max	P-values Δ	
Age (in yrs.)	269	44.55	14.87	12	84	0.72	0.24
Female (=1)	372	0.41		0	1	0.21	0.50
Height (in cm)	269	177.2	10.26	145	201	0.08	0.88
Years lived in The Netherlands	269	40.28	18.05	0.01	84	0.85	0.37
Native (=1)	269	0.82		0	1	0.70	0.76
Risk taking (Low=0, High=10)	271	6.51	1.72	0	10	0.47	0.71
Christian faith	272	0.21		0	1	0.75	0.49
Muslim faith	272	0.05		0	1	0.55	0.51
No religious faith (incl. Atheist)	272	0.63		0	1	0.41	0.38
Other faiths	272	0.11		0	1	0.20	0.98
Responded to survey	372	0.73		0	1	0.64	0.15
Acknowl. seeing glove drop (analysis sample)	372	0.83		0	1	0.41	0.68

Note: Natives are participants who have lived their entire lives in The Netherlands. Christian faith includes Catholics and Protestants. P-values stem from two-sided t-tests on the full sample.

Table 2: Sample sizes by treatment groups

	Analysis Sample			*Full Sample*		
	Time pressure	Time delay	Total	Time pressure	Time delay	Total
Control	77	77	154	92	91	183
Headscarf	81	75	156	94	95	189
Total	158	152	310	186	186	372

Note: The analysis sample excludes people who indicated that they had not seen the glove drop.

Table 3: Regression estimates

	(1)	(2)	(3)	(4)
Time delay	-0.224*		-0.225*	-0.195*
	(0.050)		(0.050)	(0.069)
Headscarf		-0.067	-0.072	-0.042
		(0.051)	(0.050)	(0.061)
Time delay * Headscarf				-0.061
				(0.100)
Constant	0.823*	0.747*	0.859*	0.844*
	(0.030)	(0.035)	(0.039)	(0.042)
Time delay=Headscarf (p-value)			0.03	0.03
N	310	310	310	310

Note: Linear regression model. Subjects who indicated that they had not seen the glove drop are excluded from the analysis. Results remain statistically similar when such subjects are included (see Table B2 in the Annex). Robust standard errors in parentheses. Symbols denote significance levels at +p < 0.1, *p < 0.05.

Table 4: Robustness of regression estimates to the inclusion of covariates

	(1)	(2)	(3)	(4)
Time delay	-0.225*	-0.176*	-0.168*	-0.133
	(0.050)	(0.058)	(0.058)	(0.084)
Headscarf	-0.072	0.018	0.006	0.038
	(0.050)	(0.057)	(0.058)	(0.069)
Time delay * Headscarf				-0.070
				(0.117)
Subject characteristics				
Age			0.002	0.002
			(0.002)	(0.002)
Female			-0.097	-0.097
			(0.081)	(0.081)
Native			-0.006	-0.012
			(0.074)	(0.075)
Risk taking			0.038*	0.037*
			(0.018)	(0.018)
Height			-0.002	-0.001
			(0.004)	(0.004)
Subject's religious faith				
No religion (excl. category)				
Christian			0.083	0.085
			(0.067)	(0.066)
Muslim			0.046	0.055
			(0.140)	(0.142)
Other			0.058	0.055
			(0.090)	(0.091)
Constant	0.859*	0.825*	0.790	0.773
	(0.039)	(0.046)	(0.755)	(0.760)
N	310	225	225	225

Note: Linear regression model. Subjects who indicated that they had not seen the glove drop are excluded from the analysis. Column 1 repeats baseline estimates from column 3 in Table 3. Sample size drops from column 1 to 2-4 due to survey non-response. See also Table B3 where we fill in missing covariates to investigate systematic non-response to the post-experimental survey. Robust standard errors in parentheses. Symbols denote significance levels at $+p < 0.1$, $*p < 0.05$.

Table 5: Heterogeneity as a function of subject's religious faith

	(1)	(2)
Time delay	-0.168*	-0.175*
	(0.058)	(0.057)
Headscarf	0.006	-0.042
	(0.058)	(0.078)
Subject's religious faith		
No religion (excl. category)		
Christian	0.083	0.095
	(0.067)	(0.094)
Muslim	0.046	-0.423*
	(0.140)	(0.198)
Other	0.058	-0.058
	(0.090)	(0.129)
Christian * Headscarf		-0.021
		(0.133)
Muslim * Headscarf		0.652*
		(0.236)
Other * Headscarf		0.233
		(0.161)
Constant	0.790	1.060
	(0.755)	(0.726)
N	225	225

Note: Linear regression model. Subjects who indicated that they had not seen the glove drop are excluded from the analysis. Other covariates, as in Table 4, are included but not shown. Robust standard errors in parentheses. Symbols denote significance levels at +p < 0.1, *p < 0.05.

8. **Photos**[17]

Photos 1 and 2: Shown in main text.

Photos 3 and 4: Time pressure condition in headscarf and control identities

Photos 5 and 6: Time delay treatment in headscarf and control identities

Note: The researcher was sitting on the bench behind the participant, permitting private and anonymous decisions by the participants.

[17] Photos are illustrative and were taken after the experiment was completed.

Photos 7 and 8: Participant characteristics

Note: To be considered for the experiment participants must be: walking alone, in no visible hurry nor distracted, and with no other subject close by.

Photo 9: Typical helping behavior: Participant picks the glove and gives it back

Photo 10: Typical helping behavior: Pointing and alerting with voice

Photo 11: Participant defects

Photo 12: Survey after the experiment

9. Annex

A. Photos

Photo A.1: Front view of location

Photo A.2: Left view of location

Photo A.3: Right view of location

Photo A.4: Back view of location

B. Tables

Table B1: Types of helping behaviors (in %, conditional on helping)

	Total rates		Control		Headscarf	
	Control	Headscarf	Time pressure	Time delay	Time pressure	Time delay
Picking glove and giving back	52.58	41.51	48.48	58	41.54	41.46
Signals and voice alert	35.34	44.34	39.39	30	44.62	43.90
Only voice alert	12.07	14.15	12.12	12	13.85	14.63

Table B2: Comparing full vs. analysis samples (including vs. excluding subjects who indicated they had not seen the glove drop)

	(1)	(2)	(1) vs. (2) P-Value Δ
Time delay	-0.157*	-0.195*	0.37
	(0.071)	(0.068)	
Headscarf	-0.015	-0.042	0.55
	(0.067)	(0.061)	
Time delay * Headscarf	-0.103	-0.061	0.47
	(0.099)	(0.100)	
Constant	0.707*	0.844*	
	(0.048)	(0.041)	
Sample	Full	Analysis	
N	372	310	

Note: Seemingly unrelated regression model. Column 1 includes subjects who indicated they had not seen the glove drop (full sample). Column 2 is estimated on the analysis sample. The final column tests for differences in coefficients between columns 1 and 2 using a seemingly unrelated regression model. Robust standard errors in parentheses. Symbols denote significance levels at +p < 0.1, *p < 0.05.

Table B3: Robustness of estimates when filling in missing survey data

	(1)	(2)	(3)
Time delay	-0.219*	-0.217*	-0.190*
	(0.050)	(0.050)	(0.070)
Headscarf	-0.065	-0.076	-0.050
	(0.049)	(0.050)	(0.062)
Time delay * Headscarf			-0.054
			(0.100)
Non-response to survey	-0.129*	-0.130*	-0.132*
	(0.059)	(0.059)	(0.059)
Subject characteristics			
Age		0.002	0.002
		(0.002)	(0.002)
Female		-0.106+	-0.107+
		(0.063)	(0.063)
Native		0.000	-0.005
		(0.075)	(0.076)
Risk taking		0.035+	0.034+
		(0.018)	(0.018)
Height		-0.002	-0.002
		(0.004)	(0.004)
Subjects religious faith			
No religion (excl. category)		0.080	0.080
		(0.066)	(0.066)
Christian		0.074	0.081
		(0.148)	(0.149)
Muslim		0.048	0.046
		(0.092)	(0.093)
Other		-0.009	-0.010
		(0.175)	(0.177)
Constant	0.886*	0.938	0.926
	(0.042)	(0.680)	(0.682)
N	310	310	310

Note: Linear regression model. Subjects who indicated they had not seen the glove drop are excluded from the analysis. Missing covariate information was filled in with sample means in columns 2 and 3. A dummy indicator taking on a value of one for incomplete survey information (in 7 cases) and zero otherwise is included but not shown. Robust standard errors in parentheses. Symbols denote significance levels at +p < 0.1, *p < 0.05.

C. Post-experimental survey

1. Did you see me dropping the glove ?
   0. No _____. 1. Yes _____.
2. What is your age? _____ years.
3. How long have you lived in The Netherlands? _____ (in years / months).
4. Where you afraid to help (or considered it)?
   0. No _____. 1. Yes _____.
5. How difficult was it to make the decision?
   0. Not at all _____. 1. Just a little _____. 2. Quite _____. 3. A lot _____.
6. From 0 to 10, how willing are you to take risks in general? (max. is 10): Rate: _____.
7. What is your height? _____ cm.
8. What is your religion?
   1. Christian Catholic ___ 2. Christian Protestant _____ 3. Muslim _____ 4. No religious faith (incl. Atheist)_____
   5. Other _____