

**Racial Bias in the Sharing Economy and the Role of Trust and Self-Congruence**

Katrine Berg Nødtvedt<sup>1</sup>, Hallgeir Sjøstad<sup>1</sup>, Siv Rosendahl Skard<sup>1</sup>, Helge Thorbjørnsen<sup>1</sup>, Jay J.

Van Bavel<sup>2</sup>

<sup>1</sup> Department of Strategy and Management, Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway

<sup>2</sup> Social Identity and Morality Laboratory, Department of Psychology & Center for Neural Science, New York University, 6 Washington Place, NY 10003

**Author note**

The authors declare that they had no conflicts of interest with respect to their authorship or the publication of this article.

Correspondence concerning this article should be addressed to Katrine Berg Nødtvedt, Helleveien 30, 5045 Bergen, Norway, [katrine.nodtvedt@gmail.com](mailto:katrine.nodtvedt@gmail.com).

Findings from this paper were presented at the Frontiers in Service Conference in Austin, Texas, in September 2018. A preprint version of this article has been posted to the SSRN website: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3434463](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3434463).

### **Abstract**

The rise of peer-to-peer platforms has represented one of the major economic and societal developments observed in the last decade. We investigated whether people engage in racial discrimination in the sharing economy, and how such discrimination might be explained and mitigated. Using a set of carefully controlled experiments ( $N = 1,599$ ), including a pre-registered study on a nationally representative sample, we find causal evidence for racial discrimination. When an identical Airbnb apartment is presented with a racial out-group (vs. in-group) host, people report more negative attitudes towards the apartment, lower intentions to rent it, and are 25% less likely to choose the apartment over a standard hotel room in an incentivized choice. Reduced self-congruence with apartments owned by out-group hosts mediates these effects. Left-leaning liberals rated the out-group host as *more* trustworthy than the in-group host in non-committing judgments and hypothetical choice, but showed the same in-group preference as right-leaning conservatives when making a real choice. Thus, people may overstate their moral and political aspirations when doing so is cost-free. However, even in incentivized choice, racial discrimination disappeared when the apartment was presented with an explicit trust cue, as a visible top-rating by other consumers (5/5 stars).

*Keywords:* racial bias, sharing economy, trust, self-congruence

### **Public significance statement**

In three experiments ( $N = 1,599$ ), which included a pre-registered study on a nationally representative sample (Norway), we find causal evidence for racial discrimination against minority Airbnb hosts. When an identical Airbnb apartment was presented with a racial out-group (vs. in-group) host, people reported more negative attitudes towards the apartment, lower intentions to rent it, and were 25% less likely to choose the apartment over a standard hotel room in a real choice.

**Racial bias in the sharing economy and the role of trust and self-congruence**

The rise of peer-to-peer platforms has represented one of the major economic and societal developments observed in the last decade, typically referred to as the *sharing economy*. Each year, 730 million people stay at Airbnb apartments around the globe (Airbnb, n.d.) and over 10 billion Uber trips have been completed worldwide (Uber, 2018).

Unfortunately, there is growing evidence of racial discrimination on these platforms. Field experiments have demonstrated that guests with distinctively Black names are 16-40% less likely to be accepted by Airbnb hosts (Cui, Li, & Zhang, 2019; Edelman, Luca, & Svirsky, 2017). An observational study found that apartments belonging to Black Airbnb hosts were priced approximately 10% lower than similar listings belonging to White Airbnb hosts (Jaeger, Slegers, Evans, Stel, & Beest, 2019). These findings mirror the results from prior research showing that ethnic or racial minorities face discrimination in various markets (Bertrand & Mullainathan, 2004; Ondrich, Stricker, & Yinger, 1999), and suggest that such discrimination on peer-to-peer platforms may also be pervasive.

Discrimination in marketplace settings is a topic of high societal importance, but psychological research on the subject has been surprisingly sparse. Although economic research has provided a useful overview of the extent of discrimination in domains such as housing (Ondrich et al., 1999), labor (Bertrand & Mullainathan, 2004) and product markets (Zussman, 2013), less is known about psychological drivers and effective remedies. Field experiments, where fictitious requests are sent to real Airbnb hosts with either prototypical white- or black-sounding profile names, provide evidence of discrimination against ethnic minority Airbnb guests (Cui et al., 2019; Edelman et al., 2017). Both studies found that requests sent from profiles with black-sounding names were significantly less likely to be accepted by the hosts. However, providing reviews by previous hosts eliminated

discrimination, whereas a positive self-description written by the guests themselves did not have any impact (Cui et al., 2019).

Although previous research provided initial evidence for racial discrimination on Airbnb, the psychological process underlying these decisions has been left unexamined. Why do people act this way? The studies by Cui et al. (2019) and Edelman et al. (2017) both applied a theoretical framework from economics, namely the notion of "statistical discrimination" as opposed to "taste-based discrimination" (Guryan & Charles, 2013). The taste-based discrimination model states that some people might have a preference not to interact with members of certain social groups, and that they will be willing to pay a cost in order to discriminate against members of the disliked group (Becker, 1957). Statistical discrimination theory argues that discrimination in various transactions happens not because the discriminating party has a distaste for certain groups, but because a lack of precise knowledge about the specific individual leads to greater reliance on stereotypical, group-based information (Phelps, 1972).

What neither of the previous studies provide, however, is a test of which stereotypical beliefs and specific judgments are at work in producing racial discrimination on Airbnb. That is, the theoretical framework applied in previous studies does not predict what specific traits judgments are likely to place minority individuals in a negative light, which trait judgements that will influence consumer choice, and whether specific beliefs about the Airbnb host might have "spillover"-effects on how the rental apartment is perceived. Finally, statistical discrimination theory does not indicate whether there are certain groups of individuals who will be more or less likely to discriminate than others.

In our view, this suggests that a broader psychological perspective is needed to understand the drivers and remedies of racial discrimination. Moreover, by using controlled

experiments, hypotheses about the underlying decision process can be tested empirically. The studies by Cui et al. (2019) and Edelman et al. (2017) both employed a field experimental design in an Airbnb setting, which enables causal inference but does not easily allow for survey questions or other process measures. For that reason, these studies did not indicate whether discrimination was related to certain types of beliefs and not to others, or establish *why* externally provided information was more effective than self-provided information (Cui et al., 2019). As a natural next step in racial bias research in the sharing economy, we suggest that a proper understanding of the process driving discrimination is crucial for both psychological theory and applied interventions.

### **Theoretical framework**

In the current investigation, we apply a theoretical framework that integrates social psychological theories of prejudice and discrimination, as well as theories of identity-related consumer behavior to understand racial discrimination in the sharing economy. Specifically, we build on elements from the social identity perspective (Hornsey, 2008), theory of group-based trust (Foddy, Platow, & Yamagishi, 2009), intergroup threat theory (Stephan, Ybarra & Rios Morrison, 2009), and theories of identity and self-concept in consumer behavior (Sirgy, 1982; Escalas & Bettman, 2005, Berger & Heath, 2008). On this basis, we conducted three controlled experiments to test a set of specific hypotheses about racial discrimination on Airbnb and the psychological process underlying such discrimination.

### **In-group bias and the social identity perspective**

A vast literature in social psychology has been dedicated to the issues of group-based prejudice and discrimination. At the core of this research is the phenomenon of in-group bias. Across a wide range of outcomes, people display a tendency to favor their own group, seemingly only because they belong to it (Brewer, 1979; Dunham, 2018). The seminal

framework of Social Identity Theory (Tajfel, 1982) was built from this observation, suggesting that the mere act of categorizing people as in-group or out-group members will tend to produce in-group favoritism, even when the groups are assigned based on minimal criteria and there is no history of conflict between the groups (Tajfel, Billig, Bundy, & Flament, 1971). According to Social Identity Theory people derive parts of their identity from their attachments to different groups, and they tend to behave in ways that support a positive view of their in-groups (Hogg, 2016). This in-group bias manifests itself in a wide range of outcomes, from evaluating in-group members more favorably on positive traits (Platow, McClintock, & Liebrand, 1990), to allocating more rewards to the in-group at the cost of an out-group (Tajfel, 1970).

### **Social identity and trust**

One particularly important category of group membership for many people, is race and ethnicity (Richeson, & Sommers, 2016). When people encounter different potential hosts on the Airbnb platform, the social identity perspective suggests that people will have a systematic tendency to form more positive impressions of the racial in-group hosts than racial out-group hosts – even when other sources of information are identical. Building on this, we argue that there is one kind of trait judgement that is especially relevant to people's attitudes and willingness to rent an Airbnb apartment, and that is *trust*. Trust is key to facilitate economic exchange, since marketplace interactions often involve a combination of future uncertainty and asymmetric information between seller and buyer. On Airbnb, the host possesses more information about their apartment than the guest, and distrust in the host can lead to uncertainty on part of the guest as to whether photos and descriptions provided are actually accurate. The relative lack of formal regulation of Airbnb might further elevate the importance of mutual trust. However, people tend to rate in-group members as more trustworthy (Platow et al., 1990, Falk, Heine, & Takemura, 2014) and trust in-group members

more than out-group members based on the belief that in-group members will favor each other (Foddy et al., 2009). Further, the perceived untrustworthiness of out-groups is uniquely predictive of actual marketplace discrimination (Zussman, 2013). For these reasons, we predicted lower demand for apartments that are owned by out-group hosts than in-group hosts, and that lower trust perceptions of out-group hosts would partly explain this effect through statistical mediation.

If trust perception is a factor underlying discrimination, providing explicit trust cues may mitigate bias. Prior research has found that reputation-based information can reduce racial discrimination among Airbnb users (Cui et al., 2019). However, as the previous research has not included measures of psychological variables, there is still a lacking understanding of why, when and for whom reputation-based information is effective. The current experiments were designed to investigate those questions as well, to build a deeper understanding of racial discrimination in the sharing economy.

### **Social identity and self-object congruence**

In addition to trust perceptions, a second path through which social identity might lead to discrimination on Airbnb is through feelings of perceived congruence between oneself and the apartment (hereafter referred to as *self-object congruence*). Theory of identity-based consumer behavior states that people use products, brands and services in order to construct and communicate their own identity (Belk, 1988; Berger & Heath, 2008; Reed, Forehand, Puntoni, & Warlop, 2012). People prefer products and brands that converge with their real or desired sense of self (Sirgy, 1980; Aguirre-Rodriguez, Bosnjak, & Sirgy, 2012), and prefer products and brands used by in-groups rather than out-groups (Escalas & Bettman, 2005). In light of Social Identity Theory these preferences can be seen as ways to express attachment to the in-group, or they might reflect people's tendency to use group norms to guide their



behavior. In any case, if we conceive of an Airbnb apartment as an experiential product that an individual can choose to consume or not, we would expect people to favor an in-group host's apartment, in part, because people will experience greater self-object congruence with the apartment. Put differently, people will tend to prefer an apartment if they know it belongs to someone like themselves.

This prediction is also supported by research on sharing, as people are generally more willing to share items with people belonging to their in-groups, such as family or close friends (Hellwig, Morhart, Girardin, & Hauser, 2015). Conversely, people are often averse to share items with strangers (Haze, Delcourt, & Van Vaerenbergh, 2017), and particularly with disliked individuals (Newman, Diesendruck, & Bloom, 2011). In sum, the literatures on identity-based consumer behavior, sharing and contamination all support the prediction that perceived self-congruence will make a consumer more positive to use or consume an object.

If lower self-object congruence is a driver of racial discrimination, a possible strategy to mitigate discrimination could be to signal similarities between the out-group member and the in-group. According to the social identity perspective, the categorization of people into groups is a flexible process, and the criteria for parsing the social environment into "us" and "them" can vary across situations (Tajfel, 1970; Turner, Oakes, Haslam, & McGarty, 1994). Both highlighting multiple social identities (Crisp & Hewstone, 2007), and making a common social identity salient has been shown to be effective in some contexts of intergroup discrimination (Gaertner, Dovidio, & Bachman, 1996; Van Bavel & Cunningham, 2009). We experimentally test this explanation in the current research.

### **The moderating role of individual differences**

Although individuals from both sides of the political spectrum can display discrimination (Brandt, Reyna, Chambers, Crawford, & Wetherell, 2014), racial

discrimination against the out-group in the current experiments (non-Western immigrants) is more common among people with a conservative or right-leaning political ideology (Ceobanu & Escandell, 2010; Sidanius, Pratto, & Bobo, 1996). We therefore predict that people with a right-leaning (vs. left-leaning) political orientation will be more negative towards the Airbnb apartment with an out-group host. We also predict that political orientation is related to the degree to which people experience the hosts as trustworthy, and the degree to which people experience self-object congruence with the Airbnb apartment. Specifically, we expect that conservatives (to a larger extent than liberals) will rate the out-group host as less trustworthy than the in-group host, and rate the out-group host's apartment as less self-congruent.

Another dimension of enduring individual differences likely to affect evaluations of an in-group vs. out-group host and their apartments, is beliefs about the threat of the out-group in question. Perceived out-group threat is viewed as a central antecedent of discrimination across various theoretical perspectives (Böhm, Rusch, & Baron, 2018; Sherif, Harvey, White, Hood, & Sherif, 1954; Stephan & Stephan, 2000). Threat perceptions have been found to explain out-group hostility better than general prejudice measures, because they relate more closely to the specific emotional and behavioral response evoked by an out-group (Cottrell, Richards, & Nichols, 2010). Previous research has also found threat to be predictive of out-group distrust and out-group derogation (Voci, 2006). We predicted that participants' perceptions of an out-group as threatening to important aspects of their society would make participants more distrustful when encountering a member of that group on the Airbnb platform. We also predicted higher levels of perceived threat to increase motivation to dissociate oneself from the threatening group, which would manifest as reduced perception of self-object congruence with the out-group host's apartment. Research on symbolic contamination has shown that people devalue products they perceive to have been in contact with a source they regard as immoral (Newman et al., 2011), and we expect a similar effect to arise from perceptions of a

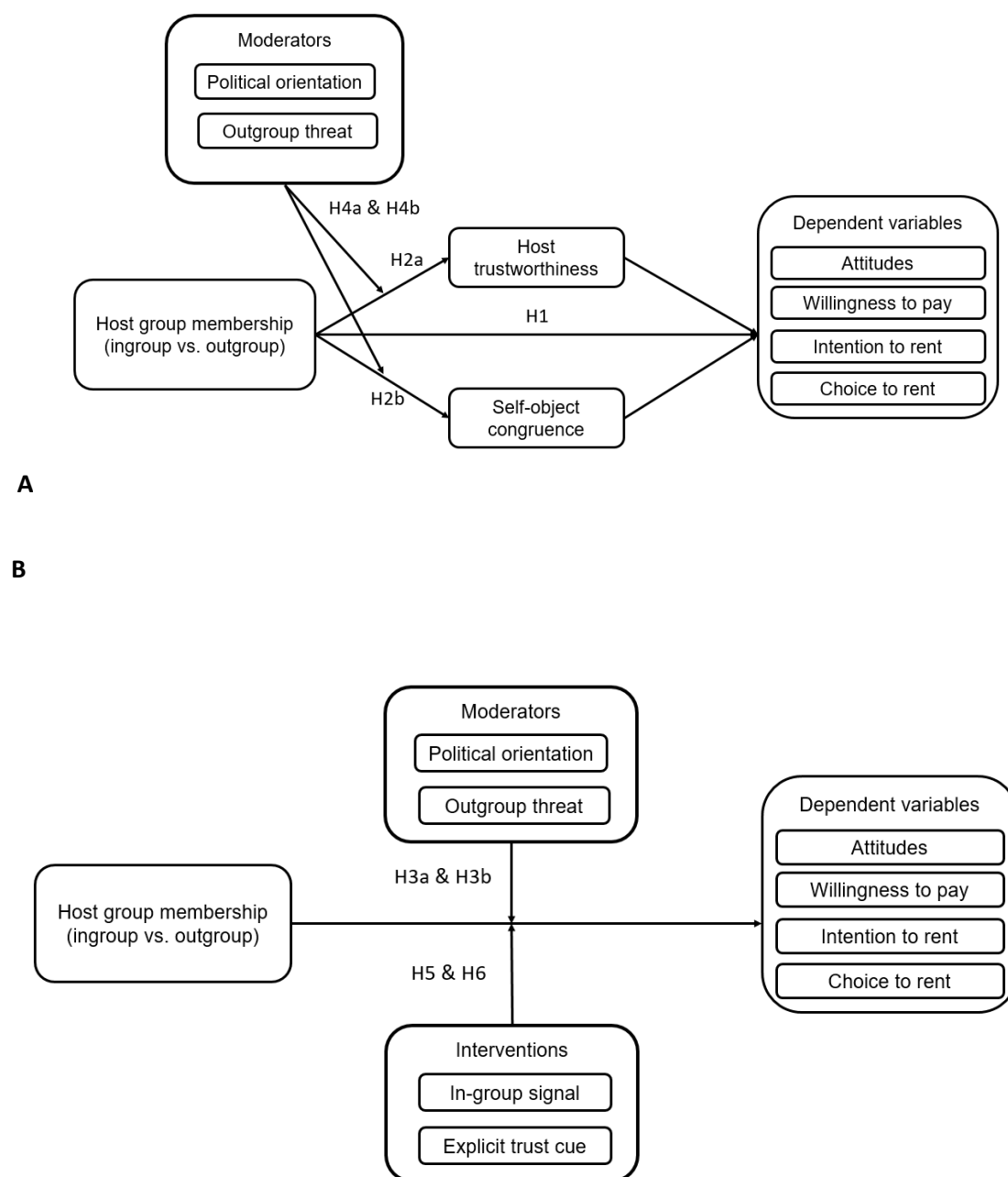
group being threatening. We expect that people who perceive an out-group as more threatening will display more reluctance towards renting an Airbnb apartment from an out-group host.

### **Current research**

The main goal of the current investigation is to build a better understanding of discrimination based on race and ethnicity in the sharing economy. In constructing our theoretical framework, we included variables from different perspectives, which made our framework more comprehensive than other models. For instance, research applying a threat-based approach to discrimination seldom measures self-congruence, and vice versa. By combining insights from modern social psychological theories of intergroup relations with insights from identity-related consumer behavior, we extend previous research on discrimination. By including measures of ideology, beliefs and social perceptions as moderator and mediator variables, we attempt to provide a more fine-grained explanation for the possibility of biased treatment of racial out-groups in this real-life marketplace setting. Crucially, we also experimentally test whether the psychology of trust can reduce racial discrimination, by testing the effect of reputation-based trust cues.

Relying on controlled experiments, we manipulate the racial group membership of the host (in-group vs. out-group) of an otherwise identical Airbnb apartment, and examine the mediating roles of self-object congruence and trust, and the moderating roles of perceived out-group threat and political orientation (see Figure 1). As outcome measures we include general evaluations of the apartment, willingness to pay, intentions to rent it, and an actual choice between the given apartment versus a standard hotel room. Crucially, we also test the effectiveness of two distinct approaches to reduce discrimination. We add an in-group signal to the profiles of out-group hosts and vary the peer ratings from previous guests to provide an

explicit trust cue. To ensure generalizability and robustness of our findings, two of our experiments use nationally representative samples of actual consumers, and our final study is a high-powered and pre-registered experiment using incentivized choice as the outcome measure.



*Figure 1:* Theoretical relationships tested in the current experiments. Panel A shows the predicted main effect of experimentally manipulating host group membership, as well as mediation effects through host trustworthiness and self-object congruence, and moderated mediation effects of political orientation and out-group threat perceptions. Panel B shows the predicted moderation effects by political orientation and out-group threat, and the predicted mitigating effects of two interventions (in-group signal and an explicit trust cue).

**Experiment 1: Investigating discrimination and effects of an in-group signal**

Experiment 1 had three main goals: First, we sought to test whether people discriminate against an out-group Airbnb host when evaluating an Airbnb apartment. Second, we aimed to test the moderating and mediating factors proposed in our theoretical framework. Third, we wished to test whether discrimination would be reduced by adding in-group signaling information to the out-group host's profile.

**Method**

**Sample.** For Experiment 1, we recruited a sample of students from a Norwegian higher education institution, through invitation by email. Sample size was based on achieving at least 80% statistical power for a one-way F test to detect a medium sized effect (Cohen's  $d = 0.5$ ), which indicated that we needed at least 159 participants. However, we put no upper restriction on participants, since a larger sample would only be desirable. We collected data over a one-week period, after which we had exceeded our sample size goal. 225 participants entered our experiment, but incomplete responses ( $n = 11$ ) were excluded from analyses (leaving a total sample of 214 participants who completed the whole experiment and were included in our analyses). The final sample consisted of 56.1% females, and the mean age was 23.7 ( $SD = 2.47$ ).

**Manipulations.** Participants were assigned to one of three different host descriptions: (1) in-group, (2) out-group or (3) out-group with in-group symbol. We manipulated the group membership of the fictitious Airbnb hosts through stated nationality, name, and picture. We chose to use a Norwegian host as the in-group host, and a non-Western immigrant host as the out-group host. The choice of non-Western immigrants as the target of discrimination reflects an attempt to operationalize racial bias in a European context. Previous investigations of ethnic discrimination in Nordic countries have often used non-Western immigrants, especially

immigrants from Muslim-majority countries (Carlsson & Eriksson, 2014; Midtbøen, 2016). In Experiment 1, the out-group host's nationality was Iraqi. The Iraqi nationality was selected because Iraqi immigrants are one of the largest groups of non-Western immigrants to Norway (Statistics Norway, 2017a), and research shows that there are negative stereotypes against this group in Norway (Bye, Herrebrøden, Hjetland, Røyset, & Westby, 2014).

The name selected for the Norwegian host was Martin (one of the top 15 most common male names in Norway, and the most popular name for 25-years old men in Norway, Statistics Norway, 2017b). The name used for the Arab host was Ahmed, the second most common Arabic male name in Norway<sup>1</sup> (Statistics Norway, 2017b). The profile pictures of the hosts were drawn from a pool of male face photos which are composites a large number of photos of individuals from different countries (The Postnational Monitor, 2011a; 2011b). The photo used for the Norwegian host was the photo for averages of European American males, since no photo has been constructed for Norwegian males. The photo used for the Iraqi host was the average photo of Iraqi males.

For the out-group host w/in-group signal, the name and photo were identical to the out-group host, but nationality was described as Norwegian-Iraqi. We also added information meant to signal affiliation and similarity with the Norwegian student sample in Experiment 1, through including information about common personal interests among students (such as an interest for travel, and outdoor sports), and a statement highlighting the host's bonds to Norway (see experimental stimuli in the supplemental materials for further details).

For all conditions, we made it clear through the apartment information that the host would not be present during the time of rental, in order to avoid potential confounds stemming from participants' expectations about in-person interaction with the host. However,

---

<sup>1</sup> The most common Arabic name in Norway is Mohammed, which we did not select because we wanted to avoid obvious connotations to Muslim religion.

previous research has found that discrimination is similarly common for shared as for non-shared apartments (Edelman et al., 2017).

**Measures.** The main dependent variables were 1) attitudes towards the Airbnb apartment, 2) intention to rent the Airbnb apartment, and 3) willingness to pay to rent the Airbnb apartment. *Attitudes* were measured with a scale composed of five items. The items were designed to tap both participants' general liking for the apartment ("How much did you like the apartment?"), their impression of attractiveness to other consumers ("How attractive do you think this apartment would be to the average student?"), and their impression of how the apartment scored on the attributes of cleanliness, standard and niceness ("Based on your general impression, how do you believe this apartment has been rated by previous guests?"). We focused the questions on attitudes towards the apartment to minimize the influence of the host and make these items conceptually similar to a decision to rent the apartment. By posing questions about participants' beliefs about attractiveness to others and perceived previous ratings by others, we aimed to reduce social desirability in responding. *Intention to rent* the Airbnb apartment was measured with a single item: "If you were to make a decision here and now, how likely is it that you would choose this apartment?" *Willingness to pay* for the Airbnb apartment was measured with a single item: "This apartment is in the price range of 500-1500 NOK [approximately \$60-180 USD] per night. How much would you be willing to pay for this apartment per night?"

We measured three items pertaining to how participants experienced the Airbnb host. One item tapped general trustworthiness perceptions: "I think [host] can be trusted." One item tapped perceived benevolence: "I think [host] is someone who first and foremost cares about what is best for his guests." The third item tapped the perceived overlap between the self and the host: "[Host] and I probably have similar values and principles." We initially conceptualized the first two items as our measure of *host trustworthiness*, and the third item

as a separate construct, labeled *self-host congruence*. However, as a part of the analysis process, we realized that there were signs of collinearity problems for these two measures. We therefore eventually chose to include the self-host congruence item as a part of the host trustworthiness measure, both because of its strong correlations with the other trust items, and because theoretically, it reflects the integrity facet of trustworthiness (Mayer, Davis, & Schoorman, 1995). See the supplemental materials for further details. In our supplemental analyses in the supplemental materials, we also present findings using both versions of the measures for full transparency. The results obtained with the different versions of the measure are almost identical, and the few discrepancies that exist do not change our main conclusions.

*Self-object congruence* was measured with one item: “I immediately felt that this apartment is ‘typically me’.” This measure was partly based on the measure of self-brand connection developed by Escalas & Bettman (2005), and was intended to capture participant’s emotional experience of overlap between their self-image and the Airbnb apartment.

*Political orientation* was measured on an 11-point scale ranging from 1 (*Left*) to 11 (*Right*). *Perceived out-group threat* was measured with two items, and referred to Muslims as the out-group: “To what extent do you think Muslims pose a threat to Norwegians?” and “To what extent do you think Muslims pose a threat to Western culture?”. The reason we chose to measure threat with reference to Muslims, and not Iraqi people, was that we expected beliefs about Muslims to be central in predicting discrimination against people from Muslim-majority countries such as Iraq. Police statistics in Norway indicate that racist and anti-Muslim speech and behavior often coincide, and that it in many cases is difficult to distinguish between these two motivations for reported hate crimes (Norwegian Police, 2019). Research also shows that stereotypes about Iraqi and other Muslim-majority country immigrants resemble stereotypes about Muslims in general (Bye et al., 2014).



All responses were recorded using 11-point Likert scales, except for willingness to pay, where responses were given as numbers in an open-ended text box. In the main text, we only present findings from variables that were applied in all three studies. For overview of all variables measured, see the measurement chapter in the supplemental materials.

**Procedure.** Participants who confirmed their voluntary participation in the experiment were randomly assigned to one of the three host conditions (in-group, out-group or out-group w/in-group signal). They were then presented with the following scenario:

Imagine that you are traveling to Copenhagen for a week-end, and are interested in renting an Airbnb apartment in the price range of 500-1500 NOK [approximately \$60-180] per night. On the next page you will be presented with an apartment in the central area of Copenhagen within this price range.

Participants were then shown a page displaying information and a photo of the Airbnb apartment (identical across all host conditions). On this page, host name and photo was also visible, and this was manipulated across conditions. After viewing the first page for at least 10 seconds, participants would click to continue to the page containing information about the host. This page displayed the name and photo of the assigned host, as well as a short text description of the host. In this text, we varied nationality of the host ("I am a [Norwegian/Iraqi/Norwegian-Iraqi] student living in Copenhagen."). For the out-group host w/in-group signal, the text contained additional information, as described in the manipulation section. Participants had to spend at least 10 seconds on this page before they could continue to the post-manipulation survey. In the post-manipulation survey, we first measured dependent variables, then mediating variables, then moderating variables and demographic/background variables. See the supplemental materials for all the stimuli used in the experiments.

**Analyses.** For mediation, moderation and moderated mediation analyses we used the PROCESS Macro (Hayes, 2018). Mediation was estimated using model 4, moderation was estimated using model 1, and moderated mediation was estimated using model 7.

## Results

We first examined whether host group affected attitudes, intentions to rent and willingness to pay by running a one-way ANOVA with the three host group conditions as independent variables. Contrary to our predictions, we found no significant main effect of host group on any of the dependent variables (attitudes:  $F(2, 211) = 1.35, p = .260, \text{partial } \eta^2 = 0.01$ , intentions:  $F(2, 211) = 1.02, p = .363, \text{partial } \eta^2 = 0.01$ , willingness to pay:  $F(2, 211) = 1.36, p = .260, \text{partial } \eta^2 = 0.01$ ). As Table 1 shows, mean scores on attitudes, intentions to rent and willingness to pay were lower for the out-group host's apartment than for the other two conditions, but none of the planned contrasts testing the mean differences between conditions were statistically significant. Neither age nor gender significantly interacted with the host group manipulation (see supplemental materials).

Table 1

*Mean scores on attitudes, willingness to pay and intentions to rent in Experiment 1*

Experimental group	Attitudes		Willingness to pay (\$)		Intentions	
	M	SD	M	SD	M	SD
Ingroup	7.12	1.46	80.85	25.20	6.17	2.19
Outgroup	6.81	1.49	74.37	23.93	5.73	2.08
Outgroup w/ingroup signal	7.19	1.44	76.45	22.90	6.24	2.58
Total	7.04	1.47	77.24	24.07	6.05	2.29

Note. M = Mean. SD = Standard deviation. None of means are significantly different in planned contrast tests.

When estimating mediation, moderation and moderated mediation effects, we conducted separate analyses contrasting two and two conditions rather than analyzing all three experimental conditions together. The main reason for this was to ease the presentation of results, since the alternative would be to use dummy coding in order to represent the three

different conditions in the same analysis. Importantly, the results and conclusions for mediation, moderation and moderated mediation analyses remain the same independent of which approach is chosen. We first present results focusing on the in-group vs. out-group contrast, before we present results involving the out-group w/in-group signal condition.

As the previous analyses had showed that there was no main effect of an out-group vs. an in-group host on the dependent variables, it was not surprising that there was not any significant indirect effects through the mediators either (see Table 2).

Table 2  
*Mediation effects of in-group vs. out-group host in Experiment 1.*

Mediator	Attitudes		Intentions		Willingness to pay	
	b	95% CI	b	95% CI	b	95% CI
Self-object congruence	-0.09	[-0.36, 0.18]	-0.11	[-0.44, 0.23]	-0.72	[-3.18, 1.48]
Host trustworthiness	0.06	[-0.10, 0.24]	0.07	[-0.13, 0.31]	0.86	[-1.57, 3.77]

Note. b = Coefficient for the indirect effect. CI = confidence interval. Results are from bootstrapped mediation analyses with 10 000 resamples. In-group host was coded as 1, out-group host as 2 in the analyses. None of the indirect effects were significant, as indicated by confidence intervals including zero.

However, analyses involving the moderating variables (political orientation and out-group threat) present an interesting picture. In order to test for potentially moderating effects of political orientation and out-group threat perceptions, we used regression analyses where we estimated the interaction effects of political orientation by host group, and out-group threat by host group. These moderation analyses revealed that the hypothesized discrimination of the out-group host was conditional on participants' political orientation and out-group threat perceptions. Specifically, out-group threat beliefs significantly moderated the effect of host in-group vs. out-group membership on participants' attitudes ( $b = -0.24$ , 95% CI [-0.45, -0.04],  $p = .023$ ), intentions to rent ( $b = -0.38$ , 95% CI [-0.68, -0.07],  $p = .015$ ) and willingness to pay<sup>2</sup> ( $b = -5.02$ , 95% CI [-8.51, -1.54],  $p = .005$ ) for the Airbnb apartment. To probe these interactions, we conducted floodlight analyses (Spiller, Fitzsimons, Lynch, & McClelland,

<sup>2</sup> All results for willingness to pay are reported in U.S. dollars.

2013), by estimating the simple effects of the independent variable (in-group vs. out-group) at all levels of the moderator. The floodlight analysis reveals a *region of significance*, which refers to the range of values of the moderator for which the simple effects of the independent variable are significant. This analysis revealed that participants with higher levels of threat responded significantly more negatively to the Airbnb apartment with an out-group (vs. in-group) host, whereas there was no significant in-group-out-group difference for participants with low threat levels. The effect of the out-group host (vs. the in-group host) was significant and negative for threat levels above 4.80 for attitudes (23.9% of the sample), above 4.55 for intentions (23.9% of the sample), and above 3.84 for willingness to pay (42.6% of the sample). Figure 2 displays this finding visually for attitudes as the dependent variable.

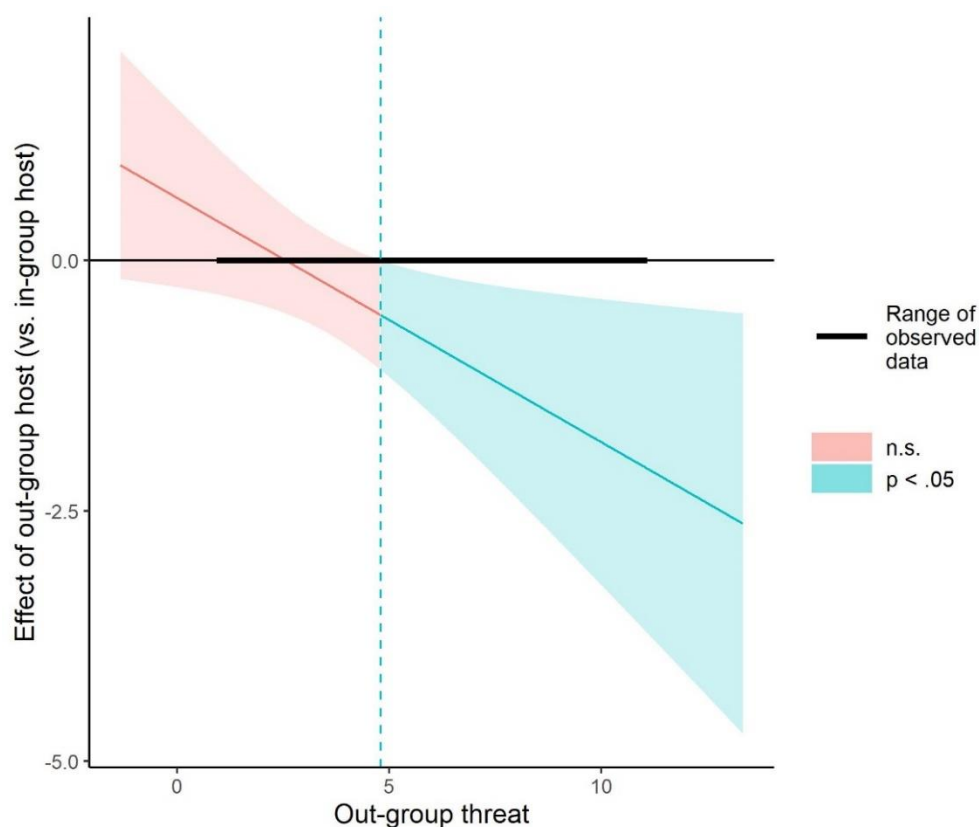


Figure 2: Effect (slope) of out-group (vs. in-group) host on attitudes towards the Airbnb apartment, showing that out-group discrimination was stronger for participants with higher levels of out-group threat. The blue area indicates the region of significance for effect of the out-group (vs. in-group) host on attitudes. Participants who were near the midpoint or higher on the threat scale displayed significant out-group discrimination. The stapled line indicates the Johnson-Neyman point, which is the point where a region of significance begins.

Out-group threat also significantly moderated the mediational effect through host trustworthiness on attitudes (95% CI [-0.19, -0.03]), intentions (95% CI [-0.25, -0.03]), and willingness to pay (95% CI [-3.52, -0.30]). Among low-threat participants, the out-group host was rated as significantly *more* trustworthy than the in-group host (*reverse discrimination*), whereas among high-threat participants, the out-group host was rated as *less* trustworthy than the in-group host (*discrimination*). This was further reflected in different mediational effects for low-threat vs. high-threat participants (see the supplemental materials for details).

Political orientation did not moderate the effect of an in-group vs. out-group host on any of the dependent variables (attitudes:  $b = -0.21$ , 95% CI [-0.45, 0.03],  $p = .091$ , intentions:  $b = -0.25$ , 95% CI [-0.60, 0.10],  $p = .167$ , willingness to pay:  $b = -1.63$ , 95% CI [-5.71, 2.44],  $p = .430$ ). However, as for out-group threat, political orientation significantly moderated the mediational effects through host trustworthiness on all the dependent variables (attitudes: 95% CI [-0.23, -0.04], intentions: 95% CI [-0.31, -0.05], willingness to pay: 95% CI [-4.14, -0.42]). Conservative participants rated the out-group host as less trustworthy, whereas liberal participants rated the out-group host as more trustworthy, which again resulted in different mediation effects for conservative vs. liberal participants (see the supplemental materials for details).

To examine the effects of the out-group host w/in-group signal, different analyses were conducted. First, we examined whether the in-group signal led to any mediational effects through host trustworthiness or self-object congruence. Results showed that participants rated the out-group host w/in-group signal as more trustworthy ( $M = 7.54$ ,  $SD = 1.71$ ) than both the out-group host ( $M = 6.77$ ,  $SD = 1.43$ ,  $t(211) = 2.79$ ,  $p = .006$ ) and the in-group host ( $M = 6.58$ ,  $SD = 1.78$ ,  $t(211) = 3.47$ ,  $p = .001$ ), and that there were positive indirect effects of the in-group signaling out-group host through host trustworthiness on all the dependent variables (see Table 3). There were however no significant effects of the out-group

host w/in-group signal on or through self-object congruence. In sum, results show that the in-group signal served to increase participants' perceived trustworthiness of the host (compared to both the in-group and the out-group host), and that this positively impacted attitudes towards the apartment and intentions and willingness to pay to rent it.

Table 3  
*Mediation effects of the out-group host w/in-group signal in Experiment 1.*

Mediator	Attitudes		Intentions		Willingness to pay	
	b	95% CI	b	95% CI	b	95% CI
Out-group host w/in-group signal vs. in-group host						
Self-object congruence	0.01	[-0.23, 0.26]	0.01	[-0.44, 0.51]	0.04	[-1.55, 2.23]
Host trustworthiness	<b>0.26</b>	<b>[0.09, 0.48]</b>	<b>0.33</b>	<b>[0.10, 0.61]</b>	<b>4.29</b>	<b>[1.31, 8.07]</b>
Out-group host w/in-group signal vs. out-group host						
Self-object congruence	0.10	[-0.16, 0.40]	0.14	[-0.23, 0.57]	0.80	[-1.16, 3.82]
Host trustworthiness	<b>0.19</b>	<b>[0.05, 0.36]</b>	<b>0.19</b>	<b>[0.01, 0.41]</b>	<b>2.54</b>	<b>[0.24, 5.26]</b>

Note. b = Coefficient for the indirect effect. CI = confidence interval. Results are from bootstrapped mediation analyses with 10 000 resamples. The out-group host w/in-group signal was coded as 2 in the analyses, and the comparison group (either in-group host or out-group host) was coded as 1. Significant effects as indicated by 95% bootstrapped confidence intervals not including zero are marked in bold.

Further, we examined how the moderators interacted with the in-group signal. Based on the finding that people with opposing political orientations and out-group threat beliefs seemed to respond differently to the out-group host, we were curious about whether the effects of the in-group signal would also be moderated by the same factors. Results were to a large extent similar for the out-group host w/in-group signal as for the out-group host. Out-group threat perceptions significantly moderated the effect of an out-group host w/in-group signal (vs. an in-group host) on attitudes ( $b = -0.31$ , 95% CI [-0.52, -0.11],  $p = .002$ ), intentions ( $b = -0.62$ , 95% CI [-0.95, -0.29],  $p < .001$ ) and willingness to pay ( $b = -5.44$ , 95% CI [-8.83, -2.06],  $p = .002$ ). Floodlight analyses revealed that for attitudes and intentions, there were two regions of significance: for threat scores below 1.57 and 1.89, there was a significant positive effect of the out-group host w/in-group signal on respectively attitudes

and intentions. For threat scores above 5.39 and 4.73, the effect on attitudes and intentions was estimated as significant and negative. The region of significance for willingness to pay was for threat scores above 3.89 (30.8% of the sample). This means that responses to the out-group host w/in-group signal (vs. the in-group host) remained significantly negative for participants with high levels of out-group threat (see Figure 3 for an illustration). However, participants with low levels of perceived out-group threat displayed *more positive* attitudes and intentions to rent the apartment presented with the in-group signaling out-group host to the apartment presented with an in-group host. Thus, for the out-group host w/in-group signal, discrimination by high-threat participants remained, but reverse discrimination by low-threat participants also occurred.

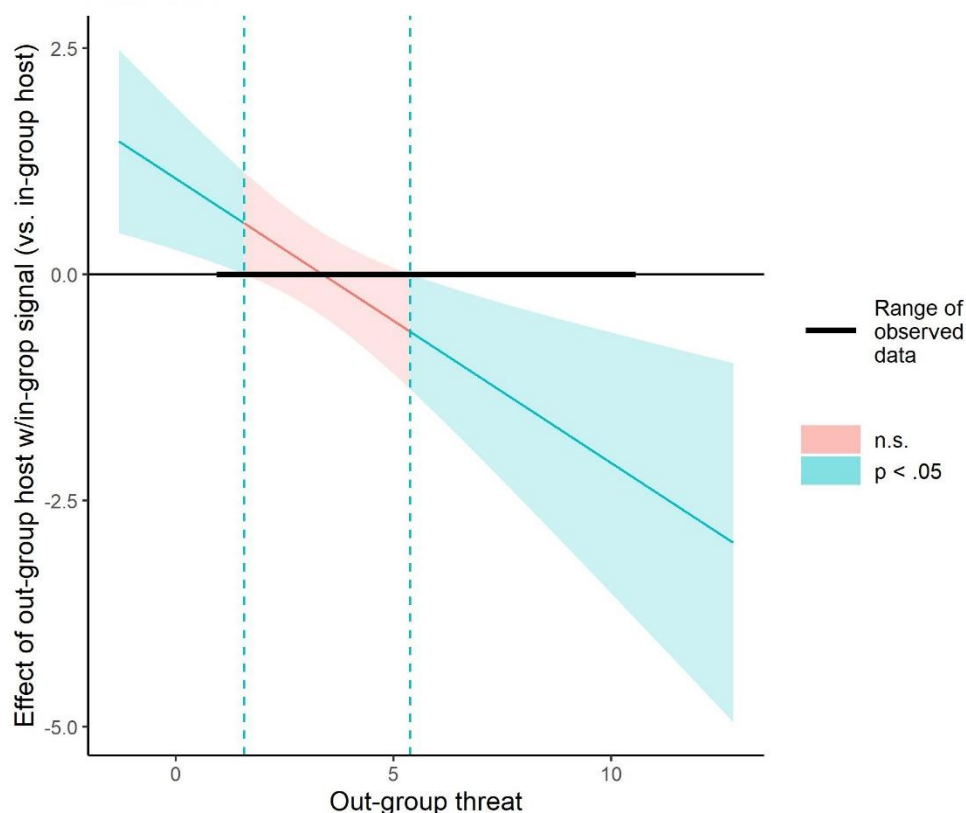


Figure 3: The conditional effect of the out-group host w/in-group signal (vs. in-group host) on attitudes towards the Airbnb apartment, showing that the out-group w/in-group signal had a negative effect when conditioning on higher threat levels, and a positive effect when conditioning on very low threat levels. The blue area indicates the region of significance for effect of the out-group (vs. in-group) host on attitudes. The stapled lines indicate the Johnson-Neyman points, which indicate the levels of the moderator where the regions of significance begin.

When comparing the out-group host with the in-group signaling out-group host, neither political orientation nor out-group threat moderated the effect of host group on the dependent variables (all  $p$ s > .05). This means that the general finding of no mean differences on the dependent variables between the out-group host and the out-group host w/in-group signal, was robust across political orientation and out-group threat perceptions.

Moderated mediation effects were also similar for the out-group host w/in-group signal and for the out-group host (see the supplemental materials for details). Liberal and low-threat participants rated the out-group host w/in-group signal as more trustworthy than the in-group host, and they rated the out-group host w/in-group signal's apartment as more self-congruent than the in-group host's apartment. For conservative and high-threat participants, these ratings were either neutral or more negative for the out-group host w/in-group signal. Thus, the in-group signal increased trustworthiness, but it did not eliminate the differences related to political orientation and out-group threat.

When comparing the out-group host w/in-group signal with the out-group host, out-group threat moderated the indirect effects through both host trustworthiness and self-object congruence. The in-group signal was more effective in increasing trustworthiness and self-object congruence among low-threat participants than among high-threat participants, resulting in the indirect effects through these variables being moderated (see supplemental materials).

Experiment 1 revealed no main effect of host race on attitudes or choice, which contradicted the previous field studies and correlational data. We did nonetheless observe discrimination against the out-group host among certain subgroups of participants, but also reverse discrimination among other subgroups. These opposing effects can contribute to explaining the lack of a main effect. Further, it seemed that adding in-group signaling information to the out-group host's profile had positive effects in terms of increasing host



trustworthiness, but that these effects did not hinder discrimination among high-threat participants.

### **Experiment 2: Conceptual replication in a large representative sample**

Although Experiment 1 revealed interesting results, the experiment relied on a non-representative student sample, which poses some limits on the generalizability of the findings (Henrich, Heine, & Norenzayan, 2010). Students tend to express less prejudice than the general population (Henry, 2008), which might have led to an underestimation of racial discrimination in Experiment 1. In Experiment 2, we therefore ran a large-scale conceptual replication of Experiment 1 on a nationally representative sample of Norwegian consumers ( $N = 584$ ) recruited through an online market research panel. The much larger, representative sample in Experiment 2 allowed us to determine which findings were robust as well as ensure they were generalizable to the national population.

### **Method**

**Sample.** For Experiment 2, we recruited participants who were members of a consumer panel run by a Norwegian market research agency. We purposely obtained a nationally representative sample in terms of gender, age and geographical location. We estimated our required sample size based on getting 80% statistical power for a one-way F test to detect a small effect of Cohen's  $d = 0.25$ . The expected effect size of  $d = 0.25$  was based on the effects observed in Experiment 1 ( $d = 0.21$ - $0.23$ ), but with a slight upward adjustment due to changes in the sample demographics and the experimental design. This power analysis led us to aim for a sample of at least 576 participants. Data was collected from 601 participants through the market research agency (only including participants who responded correctly to an initial attention check, and completed the full experiment, as in Experiment 1). Unfortunately, because of a coding error, some participants were not forced to view the manipulation pages (apartment info and host info pages) for the full length of 10

seconds. 17 participants (9 from the in-group condition, 3 from the out-group condition and 5 from the in-group-signaling out-group conditions) spent less than 7 seconds on either of these pages, and were therefore excluded from analyses<sup>3</sup>.

The final sample consisted of 584 participants. The mean age was 50.13 ( $SD = 16.41$ ), and 52.1% were female. Our sample closely matched the general Norwegian population in terms of age distribution and geographical location (see the supplemental materials for details).

**Manipulations.** As in Experiment 1, we randomly assigned participants to one of three different host descriptions: (1) in-group, (2) out-group or (3) out-group with in-group symbol. The in-group host was described as Norwegian, and we used the same name (Martin) as in Experiment 1. The picture for Martin's profile was a Stockphoto image of a Scandinavian man.

Both the out-group host and the out-group host w/in-group signal were described as Norwegian-Somali (named Abdi, photo displaying a Somali man). The reason we switched from an Iraqi immigrant to Somali immigrant as the out-group host was in order to increase the potency of our manipulation, based on knowledge that in Norway, attitudes towards Somali immigrants are somewhat more negative than attitudes towards Iraqi immigrants (Bye et al., 2014). Further, we chose to use the mixed nationality in both these conditions in order to avoid large differences in beliefs about socio-economic status of the hosts<sup>4</sup>.

---

<sup>3</sup> We judged 7 seconds as the minimum time that participants would need in order to read through the information on the Airbnb ad, and therefore the minimum time necessary to be able to count participants as sufficiently exposed to the experimental stimuli. Follow-up analyses including the 17 participants who fell below this time limit produce results that are almost identical to the results when excluding these participants. Two divergent findings exist, and are reported in the supplemental materials.

<sup>4</sup> By using a mixed-nationality target as an out-group host in Experiment 2 and 3, we made our test of discrimination in these experiments more conservative, as a clear national out-group would be more likely to evoke discrimination.

In Experiment 2, we also varied the in-group signal in a more controlled manner compared to in Experiment 1. In Experiment 1, the in-group signal consisted of both a mixed nationality, information about personal preferences and hobbies, as well as a stated attachment to the in-group country. This made the in-group signal condition inequivalent to the other two conditions, because it provided not just different information, but more information. In Experiment 2 we only varied whether the out-group host expressed attachment to the out-group (Somalis) or to the in-group (Norwegians) through the following statement in the text description of the host: “I am renting out my apartment as I frequently travel to [Somalia/Norway] to see my friends and family”.

**Measures.** All the dependent, mediating and moderating variables were measured in the same way as in Experiment 1, except for perceived out-group threat. In Experiment 1, we measured out-group threat with respect to Muslims. This was based on an assumption that participants would apply stereotypes towards Muslims to their judgements of the out-group host. However, we could not be sure that participants actually regarded the out-group host as Muslim. In Experiment 2, we decided to include items that tapped both perceived threat of Muslims, and perceived threat of Somalis in our threat measure, in order to avoid this potential limitation. We used the same items as in Experiment 1 for both these groups. See the measurement chapter in the supplemental materials for an overview of all variables measured.

**Procedure.** The procedure was identical to the one used in Experiment 1, with one addition. In order to ensure that participants attended sufficiently to the experimental instructions, we included an attention check question at the very start of the experiment, and screened out participants who failed this check. The attention check consisted of a question asking “Which of the following sports interest you the most?”, but where instructions

indicated that participants should select a specific option in order to confirm they had read the instructions.

## Results

Table 4 displays the mean scores on attitudes towards the apartment, willingness to pay, and intentions to rent, for the three host group conditions. Results from one-way analyses of variance comparing participants presented with the three different hosts revealed significant differences across the groups for attitudes towards the apartment ( $F(2, 581) = 7.42$ ,  $p = .001$ ,  $partial \eta^2 = 0.03$ ) and intentions to rent it ( $F(2, 581) = 6.13$ ,  $p = .002$ ,  $partial \eta^2 = 0.02$ ). Planned comparisons indicated that participants reported significantly more positive attitudes ( $t(581) = 3.12$ ,  $p = .002$ ,  $d = 0.31$ , 95% CI [0.11, 0.51]) and intentions ( $t(581) = 3.03$ ,  $p = .003$ ,  $d = 0.30$ , 95% CI [0.10, 0.50]) for the Airbnb apartment presented with an in-group (vs. out-group) host. For willingness to pay, the one-way ANOVA was not significant ( $F(2, 581) = 1.61$ ,  $partial \eta^2 = 0.01$ ,  $p = .200$ ), but planned contrast analysis revealed a significant in-group vs. out-group difference ( $t(376.07) = 2.32$ ,  $p = .021^5$ ,  $d = 0.24$ , 95% CI [0.04, 0.44]).

Table 4

*Mean scores on attitudes, willingness to pay and intentions to rent in Experiment 2*

Experimental group	Attitudes		Willingness to pay (\$)		Intentions	
	M	SD	M	SD	M	SD
Ingroup	6.96 <sub>a</sub>	1.69	78.71 <sub>a</sub>	30.27	5.63 <sub>a</sub>	2.72
Outgroup	6.41 <sub>b</sub>	1.82	72.02 <sub>b</sub>	26.51	4.81 <sub>b</sub>	2.69
Outgroup w/ingroup signal	6.34 <sub>b</sub>	1.75	78.52 <sub>ab</sub>	60.46	4.81 <sub>b</sub>	2.60
Total	6.57	1.77	76.39	42.04	5.08	2.69

Note. M = Mean. SD = Standard deviation. Means that do not share any of the same subscripts are significantly different ( $p < .05$ ) according to planned contrast tests.

For attitudes and intentions to rent, there were no moderating effects of gender or age on discrimination, but for willingness to pay there was a significant interaction between the in-group-out-group manipulation and age. Floodlight analyses showed that older participants

<sup>5</sup> For willingness to pay, there was significant differences in variances across groups (Levene = 5.56,  $p = .004$ ), and therefore, equal variance of the groups was not assumed in this planned comparison test.

(above 46.9 years old, 58.7% of the sample) reported significantly lower willingness to pay for the out-group host's apartment, whereas younger participants (below 46.9 years old) did not differ in their willingness to pay for the out-group vs. in-group host's apartment.

Together, these results from a large, nationally representative sample are consistent with prior fieldwork and correlational studies finding evidence of racial discrimination in peer-to-peer interactions.

Self-object congruence significantly mediated the effect of the out-group (vs. in-group) host on attitudes, intentions, and willingness to pay, but there was no significant mediation through host trustworthiness (see Table 5). This means that people reported lower levels of self-object congruence with the out-group host's Airbnb apartment, and this partially explained participants' reduced attitudes, intentions and willingness to pay for this apartment, whereas for host trustworthiness, there was no significant mediation pattern for the whole sample.

Table 5  
*Mediation effects of in-group vs. out-group host in Experiment 2.*

Mediator	Attitudes		Intentions		Willingness to pay	
	b	95% CI	b	95% CI	b	95% CI
Self-object congruence	<b>-0.19</b>	<b>[-0.36, -0.03]</b>	<b>-0.36</b>	<b>[-0.66, -0.07]</b>	<b>-2.42</b>	<b>[-4.71, -0.38]</b>
Host trustworthiness	-0.01	[-0.15, 0.12]	-0.01	[-0.13, 0.11]	-0.12	[-1.58, 1.37]

Note. b = Coefficient for the indirect effect. CI = confidence interval. Results are from bootstrapped mediation analyses with 10 000 resamples. In-group host was coded as 1, out-group host as 2 in the analyses. Significant effects as indicated by 95% bootstrapped confidence intervals not including zero are marked in bold.

Political orientation moderated the negative effect of an out-group (vs. in-group) host on the dependent variables (attitudes:  $b = -0.14$ , 95% CI [-0.29, -0.001],  $p = .048$ , intentions:  $b = -0.29$ , 95% CI [-0.51, -0.07],  $p = .009$ , willingness to pay:  $b = -3.27$ , 95% CI [-5.56, -0.98],  $p = .005$ ). Floodlight analyses indicated that the effect of the out-group (vs. in-group) host was significantly negative for moderate and conservative participants, and not significant for liberal participants (see Figure 4 for illustration and supplemental materials for details). The regions of significance began from political orientation scores above 4.48 for attitudes (68.8%

of the sample), above 4.92 for intentions (68.8% of the sample), and above 5.35 for willingness to pay (58.8% of the sample).

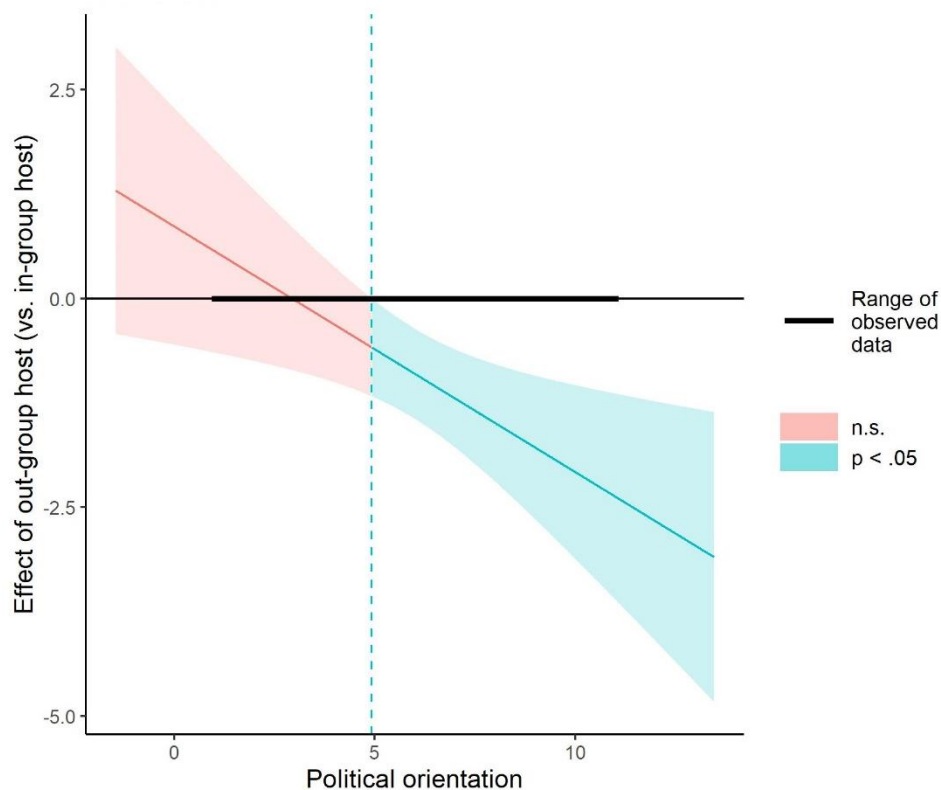


Figure 4: The conditional effect of an out-group (vs. in-group) host on intentions to rent the Airbnb apartment (Experiment 2), showing that the region of significance for the negative effect of the out-group host on intentions is for scores from 4.92 and above. Political orientation ranges from 1 (Left) to 11 (Right).

The four items used to measure out-group threat were highly correlated, and displayed almost identical relationships with the other variables in the dataset, and we therefore combined these items into a single out-group threat scale (Cronbach's  $\alpha = .97$ ). Out-group threat beliefs did not significantly moderate the effect of host group membership on the dependent variables (attitudes:  $b = -0.07$ , 95% CI  $[-0.18, 0.05]$ ,  $p = .238$ , intentions:  $b = -0.11$ , 95% CI  $[-0.29, 0.07]$ ,  $p = .230$ , willingness to pay:  $b = -1.38$ , 95% CI  $[-3.24, 0.48]$ ,  $p = .146$ ).

The moderated mediation effects in Experiment 2 followed the same pattern as in Experiment 1 (see the supplemental materials for details). The indirect effects were more negative for high-threat and conservative participants, and more positive for low-threat and liberal participants. For instance, political orientation significantly moderated the mediational

effects through host trustworthiness on attitudes (95% CI [-.18, -.05]), intentions (95% CI [-.18, -.04]), and willingness to pay (95% CI [-2.05, -.45]). Among conservative participants, the out-group host was rated as less trustworthy, which led to a negative indirect effect through trustworthiness on the dependent variables. Conversely, among liberal participants, the out-group host was rated as more trustworthy, which led to a positive indirect effect through trustworthiness (another reverse discrimination effect).

The in-group signal had mixed effects in terms of reducing discrimination in Experiment 2. Participants expressed less positive attitudes ( $t(581) = 3.53, p < .001, d = 0.36, [0.16, 0.56]$ ) and intentions ( $t(581) = 3.05, p = .002, d = 0.31, [0.11, 0.51]$ ) for the out-group host w/in-group signal's apartment than the in-group host's apartment. However, participants' willingness to pay for the Airbnb apartment of the in-group host and the out-group host w/in-group signal did not differ ( $t(288.63) = 0.04, p = .968, d = .004, [-0.20, 0.20]$ ). The presence of an extreme outlier (more than three standard deviations from the mean) in the in-group signaling outgroup condition prohibits a clear interpretation of this result, but even when recoding this extreme score to the highest score within 3 SDs from the mean, the mean difference remains insignificant ( $t(378.39) = 1.14, p = .253$ ).

In contrast to in Experiment 1, the out-group host w/in-group signal was not rated as more trustworthy than the other hosts in Experiment 2, and there were therefore no positive indirect effects through trustworthiness (see Table 6). As for the baseline out-group host, there was a negative indirect effect of the out-group host w/in-group signal (vs. the in-group host) through self-object congruence. In moderation analyses including the out-group host w/in-group signal vs. the in-group host, there was no significant reverse discrimination by liberal or low-threat participants of the out-group host w/in-group signal. However, out-group threat moderated responses to the out-group host w/in-group signal in terms of attitudes ( $b = -0.13, 95\% \text{ CI } [-0.24, -0.02], p = .018$ ) and intentions to rent ( $b = -0.20, 95\% \text{ CI } [-0.37, -0.03]$ ),

$p = .022$ ). A floodlight analysis showed that high-threat participants displayed more negative attitudes and intentions than low-threat participants. For attitudes, the region of significance started at threat scores above 3.84 (65.4% of the sample) and for intentions, the region of significance started for scores above 4.35, (61.0% of the sample). Political orientation did not significantly moderate the effect of an out-group host w/in-group signal (vs. an in-group host) on the dependent variables (all  $ps$  for interaction effect  $> .05$ ). Thus, whereas for the baseline out-group host, it was political orientation that significantly moderated discrimination, for the out-group host w/in-group signal, it was out-group threat that emerged as a significant moderator.

Table 6  
*Mediation effects of the out-group host w/in-group signal in Experiment 2.*

Mediator	Attitudes		Intentions		Willingness to pay	
	b	95% CI	b	95% CI	b	95% CI
Out-group host w/in-group signal vs. in-group host						
Self-object congruence	<b>-0.21</b>	<b>[-0.35, -0.06]</b>	<b>-0.42</b>	<b>[-0.74, -0.12]</b>	<b>-3.65</b>	<b>[-6.70, -1.03]</b>
Host trustworthiness	-0.01	[-0.17, 0.15]	-0.01	[-0.14, 0.12]	-0.06	[-1.53, 1.26]
Out-group host w/in-group signal vs. out-group host						
Self-object congruence	0.03	[-0.12, 0.17]	0.06	[-0.23, 0.36]	0.49	[-1.95, 3.00]
Host trustworthiness	-0.01	[-0.17, 0.16]	-0.004	[-0.12, 0.12]	-0.03	[-1.12, 1.20]

Note. b = Coefficient for the indirect effect. CI = confidence interval. Results are from bootstrapped mediation analyses with 10 000 resamples. The out-group host w/in-group signal was coded as 2 in the analyses, and the comparison group (either in-group host or out-group host) was coded as 1. Significant effects as indicated by 95% bootstrapped confidence intervals not including zero are marked in bold.

As was the case in Experiment 1, none of the moderators significantly interacted with the effect of the out-group host w/in-group signal vs. the out-group host (all  $ps$  for interaction effects  $> .05$ ). That means that the overall pattern was that the two out-group hosts were not treated significantly differently, and that this pattern held across different political orientations and out-group threat levels.



In terms of moderated mediation effects, the findings were similar for the out-group host w/in-group signal as for the out-group host (see the supplemental materials). Taken together, the in-group signal did not have clear mitigating effects on discrimination in Experiment 2. Discrimination remained on two of three main dependent variables, and there was no positive indirect effect of the in-group signal through increased trustworthiness ratings. A possible explanation for these findings could be that the in-group signal in Experiment 2 was more subtle than in Experiment 1. In Experiment 1, the in-group signal consisted of both a mixed nationality and information about hobbies and interests. In Experiment 2, the in-group signal was operationalized as a statement about traveling frequently to Norway. We conclude that this signal of in-group affiliation was not sufficient to reduce discrimination.

### **Experiment 3: Trust cues counteract racial discrimination**

In Experiment 3, we sought to extend the findings from the first two experiments using a realistic and incentivized choice of Airbnb apartments, and to test whether a different type of intervention could reduce discrimination. Participants in Experiment 3 were presented with a real choice between staying at an Airbnb apartment and a hotel room, if they should be the lucky winner of a lottery among the study participants. The previous studies provided initial evidence of racial discrimination, but it was on attitudinal measures, which are only modestly related to actual behavior (see Kraus, 1995). With a real choice dependent variable, we sought to get a better estimate of economic behavior and expected to reduce the amount of socially desirable responding that might be driving the *reverse discrimination* by left-wing and low-threat participants in our prior experiments.

Experiment 1 and 2 showed that the mitigating effects of in-group-signaling information were mixed. In Experiment 3, we therefore tested whether more direct, reputation-based information would be effective, by varying the presence and level of star

ratings presented with the Airbnb apartment. Experiment 3 was the largest of our studies, it included incentivized choice, and the analyses were pre-registered. As such, we have the highest confidence in the findings from this study.

## Method

Before starting data collection for Experiment 3, we pre-registered all hypotheses, measures and analyses: <https://osf.io/n8k6b/>

**Sample.** In Experiment 3, we recruited a nationally representative sample of Norwegian consumers from the same online consumer panel as used in Experiment 2<sup>6</sup> (49.6% females,  $M_{age} = 49.23$ ,  $SD = 16.95$ ). We calculated that for the current experiment, with a  $2 \times 3$  design, for an effect size of  $d = 0.28^7$ , in order to achieve at least 90% power with an alpha of 0.05 for an  $F$ -test for both main effects and interactions, we would need 649 participants ( $\approx 108$  per cell). Since the experiment included a dichotomous dependent variable (choice) with an unknown effect size, we wanted to increase sample size as much as our budget allowed, and we therefore aimed to recruit data from a sample of 800 individuals ( $\approx 133$  per cell). As in the previous experiments, we only included participants who were not screened out in the initial attention check, and who completed the full post-manipulation survey (this exclusion criteria was also pre-registered).

The final sample consisted of 801 participants. The mean age was 49.23 ( $SD = 16.95$ ), and 49.6% were female. Our sample closely matched the general Norwegian population in terms of age distribution and geographical location (see the supplemental materials for details).

**Manipulations.** Experiment 3 had a  $2$  (host group: in-group vs. out-group)  $\times 3$  (apartment rating: no rating vs. mediocre rating vs. top rating) between-subjects design. Host

---

<sup>6</sup> People who participated in Experiment 2 were not invited to participate in Experiment 3.

<sup>7</sup> The average effect size in the previous experiments.

group (in-group vs. out-group) was manipulated similarly as in the previous experiments. We again described the in-group host as a Norwegian male named Martin, and the out-group host as a Norwegian-Somali male named Abdi. In Experiment 3, we used better controlled visual stimuli for the host pictures. We selected pictures from the Chicago face database (Ma, Correll, & Wittenbrink, 2015, filename CDF-BM-029-024-N for the out-group host, and filename CDF-WM-203-023-N for the in-group host), which allowed us to match the pictures of the in-group and out-group hosts in terms of pre-rated attractiveness, threateningness, trustworthiness, and anger. Specifically, we made sure the differences in ratings for these traits was no more than 0.5 scale point on a 7-point Likert scale. This approach was chosen, as the rating data does not contain standard deviations for the individual pictures' ratings, which precluded statistical tests of differences. Rating data for the pictures are available at <https://chicagofaces.org/default/>.

Apartment ratings were manipulated by presenting a visual star rating and a corresponding number. For the mediocre rating condition, 3.5 stars were displayed, for the top rating condition, 5 stars were displayed, and for the no ratings condition, we displayed a statement saying "This property has not yet received any reviews." We confirmed through pre-testing that the 3.5 star rating was perceived by most people to be a mediocre or only slightly good score<sup>8</sup>. The hotel room option (the other option participants could choose, apart from the Airbnb apartment) was presented with a mediocre rating (3.5 stars), which was

---

<sup>8</sup> Two pre-tests confirmed this: One used a convenience sample recruited online (N = 83, Mean age = 29.6), and one used a more diverse sample recruited from a mall location (N = 24, Mean age = 48.3). 3.5 stars is actually an uncommonly low score on the real Airbnb platform, but most Norwegian consumers have no or very little experience with Airbnb, and we therefore calibrated our experiment for a sample who would not be familiar with the distribution of scores on the real platform.

constant across all conditions. The reason we presented the hotel room with a mediocre rating was to avoid floor effects in Airbnb choice. Since most consumers prefer hotels to Airbnb, we expected that a hotel with a top rating would attract a large majority of choices. Because participants were told they could win their choice of accommodation as part of a lottery prize, we also wanted to avoid giving the hotel an unrealistically low rating. Therefore, we kept the hotel room rating constant at a mediocre level.

**Measures.** The dependent, mediating and moderating variables were measured largely as in Experiment 1 and 2, with three exceptions:

In Experiment 3 we included an incentivized choice measure as a dependent variable. Specifically, we informed participants that by completing the experiment, they would enter a lottery where they could win a week-end trip to London for two people, including flights and accommodation. We then presented participants with one Airbnb apartment and one hotel room, and asked them to choose the accommodation option they would like to be included if they were to win the trip.

We modified one of the items in the host trustworthiness scale to the following: “I believe I have a lot in common with [host].”

For the perceived out-group threat measure, we decided to only refer to Somalis, since responses in Experiment 2 were very similar to questions about Muslims and questions about Somalis. We also included two questions designed to tap the dimension of realistic threat (e.g. “To what degree do you think Somali people pose a threat to the Norwegian economy?”), in addition to the symbolic threat items we had previously used (e.g. “To what extent do you think Somali people pose a threat to Western culture?”). Previous research on out-group threat has found that symbolic and realistic threat can have different effects on prejudice and

discrimination (Stephan et al., 2009). We therefore wanted to include both dimensions in our measure in order to make sure it reflected these two main subtypes of out-group threat.

**Procedure.** We applied the same attention check screening procedure as in Experiment 2. Participants were randomly assigned to one out of six experimental conditions, and were first presented with the incentivized choice measure (choosing accommodation for the trip they might win). After making a choice, they were asked to report their attitudes and willingness to pay as in the previous experiments. We then measured mediating and moderating variables. In the survey, we also asked some filler questions about attitudes and willingness to pay for the hotel room, in order to reduce demand effects. Upon completing the post-manipulation survey, we debriefed participants about the real prize of the lottery, which was an open travel voucher worth the same as trip presented in the experiment (weekend in London for two).

## Results

The critical dependent measure in this experiment was incentivized choice. Results revealed that participants chose the Airbnb apartment (vs. hotel) significantly more often when the Airbnb host was an in-group member (38.4%, 95 % CI [33.6, 43.1]) compared to when the host was an out-group member (28.9%, 95% CI [24.7, 33.5],  $\chi^2(1, 801) = 7.80$ ,  $p = .005$ , proportion difference = 9.3%, 95% CI [2.8, 15.8]). That is, in relative terms, people were approximately 25% less likely to choose the Airbnb apartment when it was presented with an out-group host compared to an in-group host. There was a significant interaction between gender and the in-group vs. out-group manipulation on choosing the Airbnb apartment vs. the hotel room. Results from Chi-square tests of independence revealed that men did not choose the Airbnb presented with an out-group host less often than the Airbnb presented with an in-group host ( $\chi^2 = 0.055$ ,  $p = .825$ ), whereas women did ( $\chi^2 = 12.804$ ,  $p <$

.001). This indicated that the discrimination on this variable was driven by the women in the sample. There was no moderating effect of age on discrimination for the choice variable.

Participants also reported significantly more positive attitudes towards the Airbnb apartment with an in-group (vs. out-group) host ( $t(799) = 2.441, p = .015, d = 0.17, 95\% \text{ CI } [0.03, 0.31]$ ). Participants did not report higher willingness to pay for the apartment with the in-group (vs. out-group) host ( $t(799) = 0.169, p = .866, d = 0.01, 95\% \text{ CI } [-0.13, 0.15]$ ), but signs of a ceiling effect on this measure prevents strong interpretations of this null effect. For attitudes and willingness to pay, there was no moderating effect of neither age nor gender (see supplemental materials).

In terms of psychological mediators, self-object congruence significantly mediated the effect of host group membership on choice, attitudes, and willingness to pay, but there were no statistically significant indirect effects of host group membership through host trustworthiness (see Table 7). This replicates the mediational findings from Experiment 2 with a similar representative sample, and supports the notion that reduced self-object congruence with the Airbnb apartment can partly explain people's reduced interest in renting from an out-group host.

Table 7

*Mediation effects of in-group vs. out-group host in Experiment 3.*

Mediator	Choice		Attitudes		Willingness to pay	
	b	95% CI	b	95% CI	b	95% CI
Self-object congruence	<b>-0.19</b>	<b>[-0.36, -0.03]</b>	<b>-0.09</b>	<b>[-0.17, -0.01]</b>	<b>-1.89</b>	<b>[-3.85, -0.31]</b>
Host trustworthiness	0.01	[-0.01, 0.04]	0.03	[-0.04, 0.11]	0.40	[-0.51, 1.53]

Note. b = Coefficient for the indirect effect. CI = confidence interval. Results are from bootstrapped mediation analyses with 10 000 resamples. In-group host was coded as 1, out-group host as 2 in the analyses. Significant effects as indicated by 95% bootstrapped confidence intervals not including zero are marked in bold.

The four items measuring out-group threat were highly correlated, and were combined into a single scale (Cronbach's  $\alpha = .90$ ). Perceived out-group threat moderated the effect of host group membership on attitudes ( $b = -0.10$ , 95% CI  $[-0.20, -0.01]$ ,  $p = .027$ ), and a floodlight analysis revealed that high-threat participants displayed significantly more negative attitudes towards the out-group (vs. in-group) host's apartment, whereas there was no host group effect for low-threat participants. The region of significance started for threat scores above 2.98 (58.2% of the sample). Out-group threat did not moderate the discrimination we observed on the Airbnb vs. hotel choice ( $b = -0.02$ , 95% CI  $[-0.16, 0.12]$ ,  $p = .759$ ), or willingness to pay for the Airbnb apartment ( $b = 0.16$ , 95% CI  $[-2.80, 3.11]$ ,  $p = .918$ ). Furthermore, political orientation did not significantly moderate any of the effects of host group membership on any of the dependent variables (choice:  $b = -0.03$ , CI  $[-0.14, 0.09]$ ,  $p = .672$ , attitudes:  $b = 0.07$ , CI  $[-0.01, 0.16]$ ,  $p = .080$ , willingness to pay:  $b = 0.13$ , CI  $[-2.52, 2.78]$ ,  $p = .922$ ). There were also fewer significant moderated mediation effects in Experiment 3 (see supplemental materials). This suggests that the effects of political ideology in the previous experiments might be primarily expressive rather than shaping actual choice behavior.

Finally, we tested whether experimentally varying trust cues had an effect on discrimination. To examine whether the rating conditions moderated the in-group-out-group difference on the binary choice variable, we conducted a logistic regression analysis of the interaction between host group membership (in-group vs. out-group), and two dummy variables representing the three rating conditions. Dummy variable number 1 represented the contrast between any ratings and no ratings (mediocre & top rating vs. no ratings), and dummy variable number 2 represented the contrast between the two types of ratings (mediocre vs. top). We chose this coding scheme in order to test both whether the mere presence of ratings would have an effect, and whether the level of the ratings would have an

effect. Dummy variable number 1 indicates whether ratings are present or not, and dummy variable number 2 indicates whether ratings were mediocre or high.

Table 8

*Coding scheme for rating condition in logistic regression analysis*

Dummy variable	No ratings	3.5 stars	5 stars
Dummy 1	0.667	-0.333	-0.333
Dummy 2	0	0.5	-0.5

Note. The table displays the values used to identify the three rating conditions (no rating, 3.5 star rating and 5 star rating) in a logistic regression analysis, by using two dummy variables.

Results revealed a significant interaction between host group membership and the mediocre vs. top rating dummy ( $b = -0.79$ , 95% CI  $[-1.54, -0.04]$ ,  $p = .038$ ), which indicated that discrimination depended on the level of the ratings. The mere presence of a trust cue (comparing no ratings with the two rating conditions) did not have a significant effect on the degree of discrimination of the out-group host ( $b = 0.03$ , 95% CI  $[-1.14, 0.28]$ ,  $p = .921$ ). See Table 9 and Figure 5 for an illustration of Airbnb choice proportions in the different host and rating conditions. Additional contrast tests are presented in the supplemental materials.

Table 9

*Percentage choosing Airbnb apartment in different host group and rating conditions in Experiment 3.*

Rating condition	Ingroup	Outgroup	Difference
No ratings	36.6%	26.1%	10.5%
3.5 stars	33.8%	17.9%	15.9%*
5 stars	44.4%	42.9%	1.5%
Total	38.4%	28.9%	9.5%*

Note. \*Chi square test significant at  $p < .05$



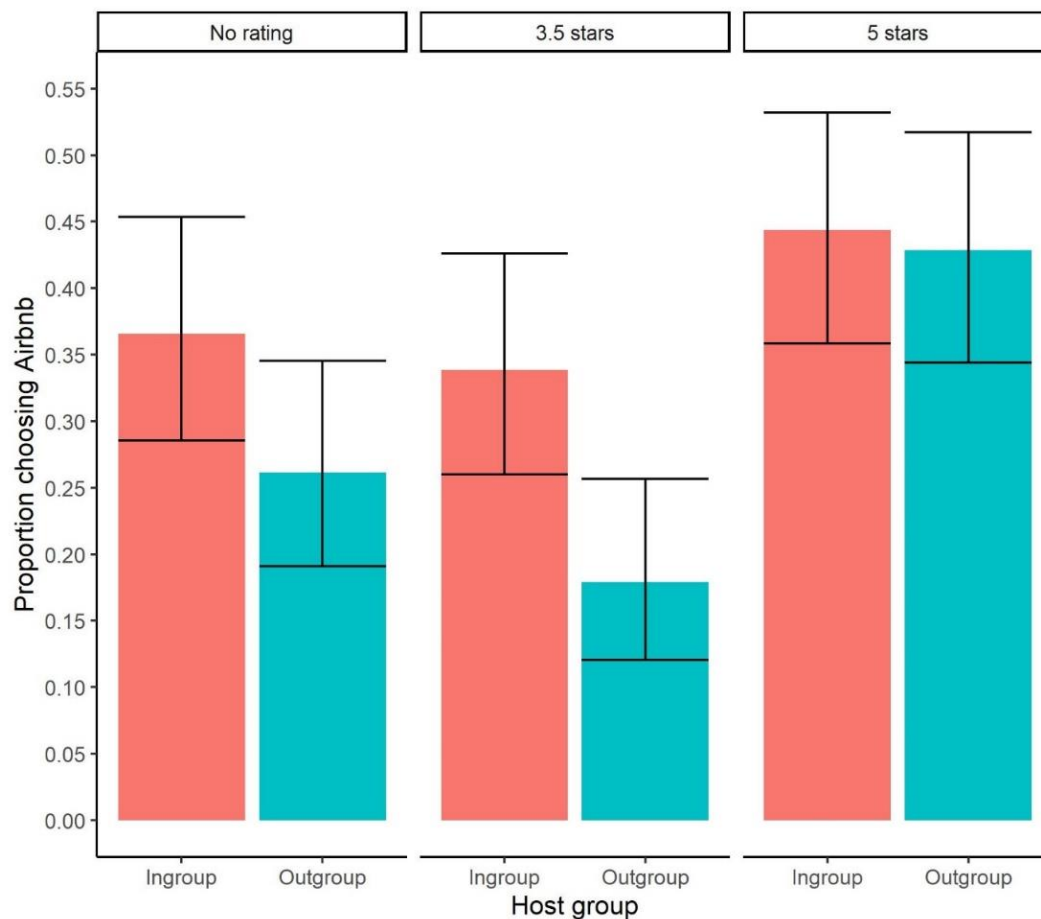


Figure 5: The figure shows the proportion of participants choosing the Airbnb as their preferred accommodation (relative to a hotel room) in a consequential choice in Experiment 3. We found evidence of discrimination in the mediocre rating condition (when the Airbnb host had 3.5 stars), but not in the top rating condition (when the Airbnb host had 5 stars). Error bars indicated 95% confidence intervals.

Probing the significant rating level  $\times$  host group interaction, we found that when the host had a mediocre rating (3.5 star), there was a statistically significant difference between the percentage of people choosing the in-group (33.8%, 95% CI [26.3, 42.2]) vs. the out-group (17.9%, 95% CI [12.3, 25.3]) Airbnb apartment (proportion difference = 15.9, 95% CI [5.4, 26.0],  $\chi^2(1, 267) = 8.83, p = .003$ ). In other words, when the Airbnb apartment had a mediocre rating, people were nearly twice as likely to choose the apartment when it was presented with an in-group host compared to when it was presented with an out-group host. However, when the Airbnb had a top (5 star) rating, there was no significant difference between the percentage choosing the in-group (44.4%, 95% CI [36.2, 52.8]) and out-group

(42.9%, 95% CI [34.8, 51.4]) Airbnb apartment (proportion difference = 1.5, 95% CI [-10.3, 13.2],  $\chi^2(1, 266) = 0.06, p = .805$ ). For the out-group host, there was also a significant difference between having a top rating vs. a mediocre rating (proportion difference = 24.9, 95% CI [14.0, 35.1],  $\chi^2(1, 267) = 19.66, p < .001$ ), whereas for the in-group host the difference between a mediocre and top rating was smaller and not statistically significant at the 0.05 alpha level (proportion difference = 10.5, 95% CI [-1.2, 21.8],  $\chi^2(1, 266) = 3.10, p = .079$ ).

Results from 2 (in-group vs. out-group)  $\times$  3 (rating condition) factorial analyses of variance further revealed that there was not a significant interaction between rating condition and host group membership in predicting neither attitudes ( $F(2, 795) = 0.32, p = .728, \text{partial } \eta^2 = .001$ ) nor willingness to pay ( $F(2, 795) = 0.93, p = .394, \text{partial } \eta^2 = .002$ ) for the Airbnb apartment. This means that the in-group-out-group differences on these variables did not change significantly across the rating conditions. See Table 10 for group means.

Table 10

*Mean scores on attitudes and willingness to pay in Experiment 3*

Experimental group	Attitudes		Willingness to pay (\$)	
	M	SD	M	SD
Ingroup total	8.60	1.42	124.74	44.70
Ingroup No rating	8.65	1.48	126.61	46.60
Ingroup 3.5 stars	8.40	1.36	124.11	44.99
Ingroup 5 stars	8.75	1.39	123.50	42.67
Outgroup total	8.34	1.62	124.17	51.90
Outgroup No rating	8.27	1.69	120.47	47.47
Outgroup 3.5 stars	8.17	1.54	123.26	61.68
Outgroup 5 stars	8.58	1.61	128.81	44.96
Total	8.47	1.52	124.46	48.41

Note. M = Mean. SD = Standard deviation. See supplemental materials for results of planned contrast tests.

### Internal meta-analysis

The central contribution in the current research, has been to gain a better understanding of the psychological mechanisms involved in racial discrimination in the sharing economy, and to test the effect of possible remedies. To provide a quantitative

overview of the simple main effect of out-group vs. in-group host on the dependent variables, we end with an internal meta-analysis. In Experiment 3, we measured real choice of an Airbnb apartment instead of intentions to rent, and these two measures are treated as a single variable in the internal meta-analysis because of their close conceptual link.

Based on Goh, Hall, & Rosenthal (2016), we conducted the internal meta-analyses using a fixed effects approach (Table 11). The meta-analytic effect sizes were statistically significant for attitudes towards the Airbnb apartment ( $d = 0.22$ ,  $p < .001$ , 95% CI [0.11, 0.32]), and for intention/choice to rent the Airbnb apartment ( $d = 0.25$ ,  $p = .002$ , 95% CI [0.14, 0.36]). For willingness to pay, the meta-analytic effect size was not statistically significant ( $d = 0.10$ ,  $p = 0.060$ , 95% CI [-0.004, 0.21]). However, signs of a ceiling effect on the willingness to pay measure in Experiment 3 prevents a strong interpretation of this result.

Table 11

*Standardized effect sizes and internal meta-analysis for the effect of an in-group (vs. out-group) Airbnb host on attitudes, intentions/choice to rent, and willingness to pay for an Airbnb apartment across three experiments (total N = 1332).*

Dependent variable	Experiment 1 (n = 143)	Experiment 2 (n = 388)	Experiment 3 (n = 801)	Meta-analytic effect size
Attitudes	$d = 0.21$	$d = 0.31^{**}$	$d = 0.17^*$	$d = 0.22^{**}$
Willingness to pay	$d = 0.26$	$d = 0.24^*$	$d = 0.01$	$d = 0.10$
Intention to rent/Real choice	$d = 0.21$	$d = 0.30^{**}$	$d = 0.23^{**}$	$d = 0.25^{**}$

Note.  $*p < .05$ ,  $**p < 0.01$ .  $d$  = Cohen's  $d$ . Internal meta-analysis conducted according to Goh et al. (2016). Intention to rent was measured in Experiment 1 and 2, whereas real choice to rent was measured in Experiment 3.

## Discussion

Across three experiments, we found that consumers discriminated against out-group hosts on Airbnb, both in terms of their attitudes and their actual choices. Most strikingly, our findings from the large-scale, nationally representative sample in Experiment 3 revealed that when an out-group (vs. in-group) host was presented together with an identical Airbnb apartment, the amount of people willing to choose that apartment in an incentivized choice dropped by 25%. Extending a previous field observation that non-White Airbnb hosts charge

prices that are approximately 10% lower than the prices charged by White Airbnb hosts for equivalent apartments (Jaeger et al., 2019), the current investigation provides causal evidence for such discrimination. Thus, people engage in costly racial discrimination towards hosts of an identical apartment.

The findings also shed light on how to reduce racial discrimination against hosts. We found strong evidence that explicit trust cues, in the form of reputation-based ratings from previous guests, can reduce such discrimination. When the Airbnb apartments were presented with either no ratings or a mediocre (3.5 star) rating, 29% and 47% fewer chose the out-group (vs. in-group) Airbnb apartment as their preferred accommodation option—even when the accommodations were otherwise identical across hosts. When the Airbnb apartments were presented with top (5 star) ratings, the in-group vs. out-group gap was almost completely eliminated (1.5% difference). This indicates that increasing the salience of top ratings for minority individuals could reduce discrimination on Airbnb and possibly other platforms in the sharing economy.

Contrary to our initial predictions, we did not find convincing evidence that highlighting points of similarity with the in-group reduce discrimination against out-group hosts in the Airbnb context. In Experiment 1, adding in-group signaling information to the out-group host's profile had positive effects in terms of increasing host trustworthiness, but these effects did not hinder discrimination among high-threat participants. In Experiment 2, with a more stringent in-group signaling manipulation, we concluded that the in-group signal did not manage to mitigate discrimination. Similarly, Cui and colleagues (2019) found that self-claimed positive information was less effective than reputation-based information in mitigating discrimination. Future research could therefore test the effects of providing signals of similarity generated by oneself vs. by trustworthy third parties.

Table 12

*Summary of findings from the three experiments*

Variable	Experiment 1 (N = 214, student sample)	Experiment 2 (N = 584, representative sample)	Experiment 3 (N = 801, representative sample)
Main effect (H1)			
Discrimination	N.s. for sample as a whole	Lower attitudes, intentions and WTP for outgroup apartment	Lower choice rate and attitudes for outgroup apartment, n.s. for WTP
Mediation (H2)			
Trust	N.s.	N.s.	N.s.
Self-object congruence	N.s.	Significant negative indirect effect for attitudes, intentions and WTP	Significant negative indirect effect for attitudes, intentions and WTP
Moderation (H3)			
Political orientation	N.s.	Significant moderation of effect of host on attitudes, intentions and WTP	N.s.
Out-group threat	Significant moderation of effect of host on attitudes, intentions and WTP	N.s.	Significant moderation of effect of host on attitudes
Moderated mediation (H4)			
Host trustworthiness	Significant moderated mediation on attitudes, intentions and WTP by both moderators	Significant moderated mediation on attitudes, intentions and WTP by both moderators	Significant moderated mediation on attitudes and WTP by out-group threat
Self-object congruence	Significant moderated mediation on attitudes, intentions and WTP by political orientation	Significant moderated mediation on attitudes, intentions and WTP by political orientation	N.s.
Mitigation interventions (H5 & H6)			
Ingroup-signaling information	The intervention increased trustworthiness ratings, but did not eliminate discrimination by high threat participants	The intervention did not increase trustworthiness ratings, and yielded mixed results in reducing overall discrimination	-
Explicit trust cues (ratings)	-	-	A top (5 star) rating significantly reduced discrimination

However, we do not interpret our findings to mean that perceived similarity is unimportant. After all, self-object congruence emerged as the most reliable mediator across our three studies. This suggests that people use host identity as a cue in forming an impression about whether an Airbnb apartment fits with their own identity, and that this judgement in turn leads to out-group host's apartments being seen as less attractive. This finding is a novel extension to the literature on racial bias, illustrating the theoretical potential of integrating consumer psychology with social psychology in order to explain behavior in marketplace contexts.

Regarding the role of political orientation and perceived out-group threat, the findings from all three experiments point to the same general pattern, although not all findings are statistically significant across the board (see the supplemental materials for illustrations). The results converged in revealing that liberal political views and low perceptions of out-group threat were related to more positive responses to the out-group host, whereas moderate and conservative political views and high perceptions of out-group threat were related to more negative responses to the out-group host. Across experiments, participants with either liberal political views or low out-group threat ratings reported *higher* trustworthiness for the out-group (vs. in-group) host, whereas participants with conservative political views or high out-group threat ratings reported *lower* trustworthiness for the out-group (vs. in-group) host. This resulted in opposite indirect effects through host trustworthiness for liberal/low-threat and conservative/high-threat participants, and explains why we do not observe a simple mediational effect through host trustworthiness in our results.

While we expected responses to the out-group host to be more negative for conservative and high-threat participants, we did not anticipate the phenomenon of reverse discrimination. One possible explanation for this behavior comes from the justification-suppression model of prejudice (Crandall & Eshleman, 2003), which states that people might

harbor negative prejudices about certain groups, but be reluctant to express them because these prejudices conflict with egalitarian values or concerns of appearing “politically correct”. In our sample, this might have been characteristic of the liberal and low-threat subgroup of our participants. Participants with liberal and pro-immigrant values might have (consciously or unconsciously) made an effort to appear un-prejudiced in their evaluations of the out-group host and his apartment, and therefore ended up rating the out-group host more favorably than the in-group host.

Interestingly, when it came to the incentivized outcome variable in Experiment 3, neither political orientation nor out-group threat beliefs mattered for participants’ decisions: People chose the in-group host’s apartment more often than the out-group host’s apartment, regardless of ideology. Whereas political orientation and out-group threat had significant moderation effects on hypothetical outcomes (attitudes and willingness to pay), we found no such effects when it came to a real and consequential choice. Additionally, in Experiment 3, political orientation and out-group threat had less impact overall, as there was no evidence for a liberal “outgroup preference” on neither incentivized choice or the evaluative mediator variables. A possible explanation for this could be that different psychological processes underlie bias on evaluative outcomes and outcomes that have real-life implications for the individual. Indeed, Dunham (2018) has made a convincing case for a distinction between in-group bias in evaluations and in-group bias in cooperative behavior. He argues that in-group bias in evaluations of others (e.g. judging in-group members to be more friendly or intelligent than out-group members) could be caused by a spill-over of positive self-regard to groups that get associated with the self, whereas in-group bias in cooperative behavior (e.g. deciding to reward in-group members more than out-group members) seems more likely to be explained by tacit norms and expectations about in-group reciprocity (Dunham, 2018). Evidence supporting this view comes from research showing that in-group bias in cooperation is

reduced when there is no interdependency of outcomes among group members, but that evaluative biases can remain (Balliet, Wu & De Dreu, 2014).

As such, our own findings are consistent with prior work, and also point to a possible extension. In Experiment 3, political orientation did not matter for discrimination in terms of actual choice, and unlike Experiment 2, it did not matter for evaluative ratings either. One way to interpret this could be that for liberal participants, choosing the hotel room over the out-group host's Airbnb apartment would have created an aversive state of cognitive dissonance in Experiment 2, when expressing one's moral aspirations was cost-free. In Experiment 3, however, economic incentivization created a practical dimension to the choice, which possibly reduced the salience or weight of moral motives. Thus, when liberal participants were placed in this real choice scenario, they showed a similar ingroup preference as conservative participants, and presumably were less bothered by it than they would have been in a purely hypothetical exercise.

More generally, our results points to the importance of distinguishing between situations where discrimination is measured on a purely evaluative level, and situations where discrimination happens in a context of potential reciprocal behavior. For instance, evaluating an Airbnb apartment can be seen as a mainly evaluative judgement, whereas actually choosing to stay in an Airbnb apartment involves cooperative aspects like enacting trust. One might speculate that motivated cognitive processes like suppression of prejudice could be more likely to affect the evaluative types of outcomes, compared to the types of outcomes with real risk to the target individual. This appears to us an interesting avenue for future research. Given that most research in social psychology is based on attitude ratings and hypothetical choices, with no measure of actual behavior (Baumeister, Vohs, & Funder, 2007; Dolinski, 2018), this discrepancy serves as a reminder of why we should combine hypothetical outcomes with incentivized choices as often as possible. Especially when social



desirability is relevant, the attitude-behavior gap is likely to occur when the person can signal their political identity and moral aspirations at no cost, possibly deceiving both themselves and others at once (von Hippel & Trivers, 2011). Systematic variation of the real cost of decision-making can therefore be leveraged to provide a better understanding of the underlying psychology, and may also improve the applied relevance to the real world.

**Generalizability and limitations.** Although our experiments focus on the specific marketplace context of Airbnb, we suggest that the findings are relevant to peer-to-peer platforms more generally, as well as other transactional contexts where people must rely on trust perceptions and judgements under uncertainty. Finally, our study findings also contribute to general theoretical knowledge about drivers and mitigation strategies that apply to racial discrimination. The degree of generalizability across different contexts and choice environments should be examined empirically in future research.

In the theoretical framework applied in this research, we attempt to strike a balance between comprehensiveness and parsimony. As a consequence, there are additional variables that we imagine could have contributed with further explanation of discrimination, that we have left unmeasured. One example is measuring feelings of threat as a mediating factor. It follows from our theoretical reasoning that when encountering an out-group host, participants might experience the host as threatening, which could result in negative evaluations and intentions of the Airbnb apartment. However, we argue that by measuring host trustworthiness, we should to a large extent be able to capture the same phenomenon, since experiencing threat could be seen as the opposite of experiencing trust. Similarly, we have chosen not to measure general trustworthiness perceptions towards the out-group, because we do measure out-group threat.

Our experiments focus on discrimination against racial minorities, specifically immigrant hosts from non-Western, Muslim majority countries, which is a group that is often the target of prejudice and discrimination in Norway (Bye et al., 2014). Thus, findings might be more relevant to discrimination of stigmatized groups as opposed to any type of intergroup context. Part of the background for selecting a negatively stereotyped group as the out-group was that our design (a hypothetical survey experiment with restrictions on maximum sample size) would not be sufficient to detect very subtle discrimination effects. In short: we aimed for a strong rather than subtle manipulation of the in-group-out-group dimension. Future research may explore whether discrimination might also arise for more minimal groups in the Airbnb context. Another suggestion for future research is to include manipulation checks at the end of the post-manipulation survey. We did not implement manipulation checks in the current research, but we acknowledge that this could have been useful.

## **Conclusion**

In conclusion, the current research provides causal evidence for racial discrimination in the sharing economy, and shows that reputation-based information can be highly effective in reducing such discrimination when real choices are made. Our findings not only reveal how racial discrimination can enter these decisions, they also offer the possibility for change. Large platforms can easily scale insights from research to reduce discrimination and promote greater fairness in the sharing economy. In an age where the economy has become decentralized, it is more important than ever to understand the individual psychology behind economic decision-making.

## **Open practices statement**

Experiment 1 and 2 were not pre-registered. The pre-registration of Experiment 3 can be accessed at <https://osf.io/n8k6b>. Data, materials and an overview of measures from all

three experiments are available at <https://osf.io/ak35s/>. We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the studies.

**Ethics statement**

We complied with all relevant ethical regulations regarding human research participants, including the guidelines from the Helsinki Declaration. As Norwegian laws and regulations does not require review by an institutional review board for non-medical, low-risk research with human participants, we did not submit the project to such a review. Informed consent was obtained from all participants.

### References

- Aguirre-Rodriguez, A., Bosnjak, M., & Sirgy, M. J. (2012). Moderators of the self-congruity effect on consumer decision-making: A meta-analysis. *Journal of Business Research*, 65, 1179–1188. <https://doi.org/10.1016/j.jbusres.2011.07.031>
- Airbnb. (n.d.). Fast facts. Retrieved June 24, 2019, from <https://press.airbnb.com/fast-facts/>
- Balliet, D., Wu, J., & De Dreu, C. K. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin*, 140, 1556–1581. <https://doi.org/10.1037/a0037737>
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, 2, 396–403. <https://doi.org/10.1111/j.1745-6916.2007.00051.x>
- Becker, G.S. (1957). *The economics of discrimination*. Chicago: University of Chicago Press.
- Belk, R. W. (1988). Possessions and the extended self. *Journal of Consumer Research*, 15, 139–168. <https://doi.org/10.1086/209154>
- Berger, J., & Heath, C. (2008). Who drives divergence? Identity signaling, outgroup dissimilarity, and the abandonment of cultural tastes. *Journal of Personality and Social Psychology*, 95, 593–607. <https://doi.org/10.1037/0022-3514.95.3.593>
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review*, 94, 991–1013. DOI: 10.1257/0002828042002561
- Böhm, R., Rusch, H., & Baron, J. (2018). The psychology of intergroup conflict: A review of theories and measures. *Journal of Economic Behavior and Organization*. <https://doi.org/10.1016/j.jebo.2018.01.020>
- Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The

ideological-conflict hypothesis: Intolerance among both liberals and conservatives.

*Current Directions in Psychological Science*, 23, 27-34.

<https://doi.org/10.1177/0963721413510932>

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, 86, 307–324.

<https://doi.org/10.1037/0033-2909.86.2.307>

Bye, H. H., Herrebrøden, H., Hjetland, G. J., Røyset, G. Ø., & Westby, L. L. (2014).

Stereotypes of Norwegian social groups. *Scandinavian Journal of Psychology*, 55, 469-476. <https://doi.org/10.1111/sjop.12141>

Carlsson, M., & Eriksson, S. (2014). Discrimination in the rental market for apartments.

*Journal of Housing Economics*, 23, 41-54. <https://doi.org/10.1016/j.jhe.2013.11.004>

Ceobanu, A. M., & Escandell, X. (2010). Comparative analyses of public attitudes toward immigrants and immigration using multinational survey data: A review of theories and research. *Annual Review of Sociology*, 36, 309–328.

<https://doi.org/10.1146/annurev.soc.012809.102651>

Cottrell, C. A., Richards, D. A. R., & Nichols, A. L. (2010). Predicting policy attitudes from general prejudice versus specific intergroup emotions. *Journal of Experimental Social Psychology*, 46, 247–254. <https://doi.org/10.1016/j.jesp.2009.10.008>

Crandall, C. S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin*, 129, 414–446.

<https://doi.org/10.1037/0033-2909.129.3.414>

Crisp, R. J., & Hewstone, M. (2007). Multiple social categorization. *Advances in*

*Experimental Social Psychology*, 39, 163-254. <https://doi.org/10.1016/S0065->

2601(06)39004-1

- Cui, R., Li, J. & Zhang, D. J. (2019). Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on Airbnb. *Management Science*. Advance online publication. <https://doi.org/10.1287/mnsc.2018.3273>
- Doliński, D. (2018). Is psychology still a science of behaviour? *Social Psychological Bulletin*, 13(2), 1-14.
- Dunham, Y. (2018). Mere membership. *Trends in Cognitive Sciences*, 22(9), 780-793. <https://doi.org/10.1016/j.tics.2018.06.004>
- Edelman, B. G., Luca, M., & Svirsky, D. (2017). Racial discrimination in the sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics*, 9(2), 1–22. <https://doi.org/10.2139/ssrn.2701902>
- Escalas, J. E., & Bettman, J. R. (2005). Self-construal, reference groups, and brand meaning. *Journal of Consumer Research*, 32, 378-389. <https://doi.org/10.1086/497549>
- Falk, C. F., Heine, S. J., & Takemura, K. (2014). Cultural variation in the minimal group effect. *Journal of Cross-Cultural Psychology*, 45, 265-281. <https://doi.org/10.1177/0022022113492892>
- Foddy, M., Platow, M. J., & Yamagishi, T. (2009). Group-based trust in strangers: The role of stereotypes and expectations. *Psychological Science*, 20, 419-422. <https://doi.org/10.1111/j.1467-9280.2009.02312.x>
- Gaertner, S. L., Dovidio, J. F., & Bachman, B. A. (1996). Revisiting the contact hypothesis: The induction of a common ingroup identity. *International Journal of Intercultural Relations*, 20, 271–290. [https://doi.org/10.1016/0147-1767\(96\)00019-3](https://doi.org/10.1016/0147-1767(96)00019-3)
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some

- arguments on why and a primer on how. *Social and Personality Psychology Compass*, 10, 535-549.
- Guryan, J., & Charles, K. K. (2013). Taste-based or statistical discrimination: The economics of discrimination returns to its roots. *The Economic Journal*, 123, 417–432.  
<https://doi.org/10.1111/econj.12080>
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Publications.
- Hazée, S., Delcourt, C., & Van Vaerenbergh, Y. (2017). Burdens of access: understanding customer barriers and barrier-attenuating practices in access-based services. *Journal of Service Research*, 20, 441-456. <https://doi.org/10.1177/1094670517712877>
- Hellwig, K., Morhart, F., Girardin, F., & Hauser, M. (2015). Exploring different types of sharing: A proposed segmentation of the market for “sharing” businesses. *Psychology & Marketing*, 32, 891-906. <https://doi.org/10.1002/mar.20825>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61-83. <https://doi.org/10.1017/S0140525X0999152X>
- Henry, P. J. (2008). College sophomores in the laboratory redux: Influences of a narrow data base on social psychology's view of the nature of prejudice. *Psychological Inquiry*, 19(2), 49-71. <https://doi.org/10.1080/10478400802049936>
- Hogg, M. A. (2016). Social Identity Theory. In S. McKeown et al. (Eds.), *Understanding Peace and Conflict Through Social Identity Theory, Peace Psychology Book Series* (pp. 3-15). Springer, Cham. DOI 10.1007/978-3-319-29869-6\_1
- Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and Personality Psychology Compass*, 2(1), 204-222.

- Jaeger, B., Slegers, W., Evans, A. M., Stel, M., & Beest, I. Van. (2019). The effects of facial attractiveness and trustworthiness in online peer-to-peer markets. *Journal of Economic Psychology*. <https://doi.org/S0167487018302794>
- Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin*, 21, 58-75.  
<https://doi.org/10.1177/0146167295211007>
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47, 1122-1135. doi: 10.3758/s13428-014-0532-5
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20, 709-734.  
<https://doi.org/10.5465/amr.1995.9508080335>
- Midtbøen, A. H. (2016). Discrimination of the second generation: Evidence from a field experiment in Norway. *Journal of International Migration and Integration*, 17(1), 253-272. DOI 10.1007/s12134-014-0406-9
- Newman, G. E., Diesendruck, G., & Bloom, P. (2011). Celebrity contagion and the value of objects. *Journal of Consumer Research*, 38, 215-228. <https://doi.org/10.1086/658999>
- Norwegian Police. Oslo Police District. (2019). *Hatkriminalitet: Anmeldt hatkriminalitet 2018* [Hate crime: Filed reports of hate crime 2018]. Retrieved from [https://www.politiet.no/globalassets/dokumenter/oslo/rapporter/anmeldt-hatkriminalitet-oslo/Anmeldt\\_hatkriminalitet\\_i\\_Oslo\\_2018](https://www.politiet.no/globalassets/dokumenter/oslo/rapporter/anmeldt-hatkriminalitet-oslo/Anmeldt_hatkriminalitet_i_Oslo_2018)
- Ondrich, J., Stricker, A., & Yinger, J. (1999). Do landlords discriminate? The incidence and causes of racial discrimination in rental housing markets. *Journal of Housing Economics*



8, 185–204. <https://doi.org/10.1006/jhec.1999.0249>

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, 62, 659–661.

Platow, M. J., McClintock, C. G., & Liebrand, W. B. (1990). Predicting intergroup fairness and ingroup bias in the minimal group paradigm. *European Journal of Social Psychology*, 20, 221–239. <https://doi.org/10.1002/ejsp.2420200304>

Reed II, A., Forehand, M. R., Puntoni, S., & Warlop, L. (2012). Identity-based consumer behavior. *International Journal of Research in Marketing*, 29, 310–321. <https://doi.org/10.1016/j.ijresmar.2012.08.002>

Richeson, J. A., & Sommers, S. R. (2016). Toward a social psychology of race and race relations for the twenty-first century. *Annual Review of Psychology*, 67, 439–463. <https://doi.org/10.1146/annurev-psych-010213-115115>

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., Sherif, C. W. (1954). *Intergroup conflict and cooperation: The Robbers Cave experiment*. Norman, Oklahoma: University Book Exchange.

Sidanius, J., Pratto, F., & Bobo, L. (1996). Racism, conservatism, Affirmative Action, and intellectual sophistication: A matter of principled conservatism or group dominance? *Journal of Personality and Social Psychology*, 70, 476–490. <https://doi.org/10.1037/0022-3514.70.3.476>

Sirgy, M. J. (1982). Self-concept in consumer behavior: A critical review. *Journal of Consumer Research*, 9, 287–300. <https://doi.org/10.1086/208924>

Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., & McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research*, 50, 277–288. <https://doi.org/10.1509/jmr.12.0420>

Statistics Norway. (2017a). Innvandrere i Norge 2017 (Immigrants in Norway 2017).

Retrieved from <https://www.ssb.no/befolkning/artikler-og-publikasjoner/innvandrere-i-norge-2017>

Statistics Norway. (2017b). Navn: Statistikkbank (Names: Statistics bank). Retrieved from

<https://www.ssb.no/befolkning/statistikker/navn>

Stephan, C. W., & Stephan, W. S. (2000). An integrated threat theory of prejudice. In S.

Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23-46). Mahwah, N.J:

Lawrence Erlbaum Associates.

Stephan, W. G., Ybarra, O., & Rios Morrison, K. (2009). Intergroup Threat Theory. In T.

Nelson (Ed.), *Handbook of prejudice* (pp. 43-59). Mahwah, NJ: Lawrence Erlbaum

Associates.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223(5), 96-

103. DOI: 10.2307/24927662

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*,

33, 1–39. <https://doi.org/10.1146/annurev.ps.33.020182.000245>

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and

intergroup behaviour. *European Journal of Social Psychology*, 1, 149-178.

<https://doi.org/10.1002/ejsp.2420010202>

The Postnational Monitor. (2011a). World of facial averages: Populations of the Americas.

[Blog post]. Retrieved from <https://pmsol3.wordpress.com/2009/10/10/world-of-averages-populations-of-the-americas/>

The Postnational Monitor. (2011b). World of facial averages: Middle Eastern, Central Asian

and South Asia. [Blog post]. Retrieved from

<https://pmsol3.wordpress.com/2011/04/07/world-of-facial-averages-middle-eastern-and-central-asian/>

Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective:

Cognition and social context. *Personality and Social Psychology Bulletin*, 20, 454-463.

<https://doi.org/10.1177/0146167294205002>

Uber. (2018, July 24). 10 billion. Retrieved from <https://www.uber.com/newsroom/10-billion/>.

Van Bavel, J. J., & Cunningham, W. A. (2009). Self-categorization with a novel mixed-race group moderates automatic social and racial biases. *Personality and Social Psychology Bulletin*, 35, 321-335. <https://doi.org/10.1177/0146167208327743>

Voci, A. (2006). The link between identification and in-group favouritism: Effects of threat to social identity and trust-related emotions. *British Journal of Social Psychology*, 45, 265-284. <https://doi.org/10.1348/014466605X52245>

Von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, 34, 1-56.

Zussman, A. (2013). Ethnic discrimination: Lessons from the Israeli online market for used cars. *The Economic Journal*, 123, 433-468. <https://doi.org/10.1111/eoj.12059>