ISSN (Print) 0473-453X
Discussion Paper No. 1317
ISSN (Online) 2435-0982

CONTACT AND THE PRICE OF PREJUDICE

Andreas Makoto Fukuda Andersen
Gwen-Jirō Clochard
Guillaume Hollard
Andreas Kotsadam

July 2026

The Institute of Social and Economic Research
The University of Osaka
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Contact and the Price of Prejudice*

Andreas Makoto Fukuda Andersen[1], Gwen-Jirō Clochard[2,3,*], Guillaume Hollard[4], and Andreas Kotsadam[1,5]

[1]University of Oslo

[2]Institute of Social and Economic Research, The University of Osaka

[3]Joint Initiative for Latin American Experimental Economics

[4]Ecole polytechnique, CREST and CNRS

[5]Frisch Centre for Economic Research

[*]Corresponding author: gjclochard@iser.osaka-u.ac.jp

July 2026

Abstract

People pay to discriminate. We replicate the price-of-prejudice paradigm of Hedegaard and Tyran (2018) in Japan and test whether brief cooperative contact with an outgroup member reduces the willingness to pay (WTP) for same-ethnicity coworkers. In a pre-registered laboratory experiment at the University of Osaka ($N = 164$

*We thank seminar participants at the NASFIK and RCFK conferences, and at ISER and the Osaka School of International Public Policy seminars for helpful comments. The experiment was approved by the IRB of the Institute of Social and Economic Research, University of Osaka (#20260104), and pre-registered at the AEA RCT Registry on 13 March 2026 (AEARCTR-0018089). We acknowledge funding from the Japan Society for the Promotion of Science (KAKENHI Grant Numbers JP25K16598, JP22K21358). Andreas Makoto Fukuda Andersen acknowledges financial support from the Research Council of Norway through grants to Project No. 341250. The authors used Claude Code and ChatGPT to assist with aspects of data analysis, programming, and manuscript editing. All outputs generated by these tools were carefully reviewed and validated by the authors. The authors remain solely responsible for the accuracy, integrity, and content of the work. Romain Tranchecoste provided excellent research assistance in running the sessions.

Japanese subjects), participants first performed a real-effort task, were then randomly paired with either a Japanese or an international student for a joint production round, and finally made incentivized binary choices between future partners whose name and productivity were revealed. Discrimination is clear in the data: one half of subjects choose the Japanese partner strictly more than 50% of the time in productivity-symmetric choices, and the average subject forgoes a productivity advantage of 2.4 envelopes, about 6% of pair earnings, to work with a Japanese rather than a foreign partner. Discrimination is also strongly price-sensitive, as in Hedegaard and Tyran (2018). The contact treatment reduces estimated WTP from 2.39 to 1.82, a 24% reduction, but the difference is statistically insignificant: the experiment validates the measurement paradigm while remaining underpowered for plausible effects of a single, brief contact intervention. In exploratory analyses we find that both men and women discriminate against women, at a price exceeding that of anti-foreign discrimination.

Keywords: contact hypothesis, discrimination, experiments, willingness-to-pay.
JEL Classification: J15, C92

1 Introduction

Prejudice has social and economic costs: stereotypes help sustain social inequality (Durante and Fiske, 2017), ethnic favoritism among investors depresses the value of firms (Hjort, Song and Yenkey, 2026), and ethnic divisions between coworkers lower team output (Hjort, 2014). Becker (1957) formalized taste-based discrimination as a preference for which decision makers are willing to pay. A literature pioneered by Hedegaard and Tyran (2018) provided a direct experimental measurement of the willingness to pay to discriminate among potential coworkers: in a Danish field experiment, more than a third of subjects chose a less productive partner with a same-ethnicity-sounding name, forgoing on average 8% of their earnings. Demand for discrimination is found to be strongly price-elastic.[1]

A separate literature, originating with Allport (1954), holds that prejudice "may be reduced by equal status contact between majority and minority groups in the pursuit of common goals" [2]. Contact interventions have received more attention than any other anti-prejudice tool (Bertrand and Duflo, 2017; Finseraas and Kotsadam, 2017; Finseraas et al., 2019; Clochard, 2025; Clochard, Hollard and Sene, 2026; Scacco and Warren, 2018; Elwert, Keller and Kotsadam, 2023; Achard et al., 2025), and recent meta-analyses conclude that contact does, on average, reduce prejudice (Pettigrew and Tropp, 2006; Paluck, Green and Green, 2019; Lowe, 2025; Clochard, 2026).

This paper connects the two literatures. To our knowledge, no experiment has tested whether contact reduces the *price* individuals are willing to pay to discriminate. We embed a randomized contact round inside the price-of-prejudice paradigm: Japanese subjects at the University of Osaka first perform an individual real-effort task (envelope stuffing, as in Hedegaard and Tyran 2018), are then randomly paired with either another Japanese student (control) or an international student (treatment) for a cooperative joint-production

[1] Bajzíková and Cingl (2023) replicated the measure in the Czech Republic.

[2] Allport (1954)'s contact hypothesis holds that intergroup contact reduces prejudice most effectively under four conditions: (1) *equal status* between the groups within the contact situation; (2) *common goals* that both groups work toward; (3) *intergroup cooperation* in pursuit of those goals, rather than competition; and (4) *support of authorities, law, or custom* that lends the contact institutional legitimacy.

round, and finally make 30 incentivized binary choices between potential coworkers for a future session, with each candidate's name (signaling nationality), productivity, and time slot displayed. The design answers two questions. First, does the willingness-to-pay measure of Hedegaard and Tyran (2018) replicate outside Denmark? Japan is an interesting test case: immigrants constitute less than 2% of the population and face documented discrimination in housing (Sugasawa and Harano, 2023), education (JOSHI and Tabata, 2021), and the labor market (Igarashi and Mugiyama, 2023). Second, does brief cooperative contact reduce the measured willingness to pay?

The answer to the first question is yes. Roughly half (45.4%) strictly prefer the Japanese partner in a majority of the productivity-symmetric choice situations. Choices respond strongly to the price of discrimination: a one-unit increase in the foreign candidate's productivity advantage raises the probability of choosing the foreign candidate by 2.8 percentage points in the linear model. The implied average willingness to pay in the control group is 2.39 units of productivity, about 6% of earnings, comparable to the 8% found by Hedegaard and Tyran (2018). A measurement strategy that elicits discrimination through fully incentivized choices, without deception and without making discrimination salient, replicates in a different country, subject pool, and outgroup.

The answer to the second question is that we cannot tell. The contact treatment shifts the estimated average willingness to pay from 2.39 to 1.82, a 24% reduction in the direction predicted by contact theory, but the difference of -0.57 has a standard error of 1.22: the 95% confidence interval spans reductions of more than 100% of the control mean as well as moderate increases. The estimate is underpowered rather than a precise null, and the intervention was light by construction (a single twelve-minute round of joint production). We can rule out effects as large as our pre-registered minimum detectable effect, a reduction of about 3.5 productivity units calibrated to a more optimistic baseline, but not effects of the size found in meta-analyses of comparable interventions. Section 5 draws out the implications for the design of future contact experiments.

Exploratory analyses turn up a finding we did not anticipate: discrimination against women, larger in magnitude than discrimination against foreigners, and exhibited by male

and female subjects alike. Approximately 75% of participants, both men and women, strictly prefer male partners. We also show that an apparent positive effect of cross-gender contact in the pooled sample is an artifact of unequal gender composition across treatment groups, a cautionary result for contact studies that do not randomize the attribute of interest.

Our contribution is therefore primarily one of measurement: we show that the price of prejudice can be measured reliably across contexts, we provide the first experimental estimates of the effect of contact on that price, and we establish a baseline (both the point estimate and the power calculation) for the larger or stronger interventions that a credible test will require.

2 Experimental design

2.1 Setting and sample

The experiment was conducted at the laboratory of the Institute of Social and Economic Research (ISER), University of Osaka, in March 2026. Japanese subjects were recruited through ORSEE, the standard online recruitment system for economic experiments (Greiner, 2015). To ensure a sufficient share of foreigners, international students were additionally recruited through the university's international student network. In total, 220 subjects participated across 20 sessions: 164 Japanese students, who form the analysis sample, and 56 international students. Sessions were planned for 12 participants (9 Japanese and 3 international); no-shows left some sessions with 10. Sessions lasted approximately 45 minutes, with an average payment of JPY 1,950 ($\approx$ USD 12.50).

Of the 164 Japanese subjects, 108 were paired with another Japanese student (control) and 56 with an international student (treatment). Among treated subjects, 36 were paired with a Chinese student, 18 with an Indian student, and 2 with a Ugandan student (Table A.1).

2.2 Design

The experiment proceeded in three stages, following Hedegaard and Tyran (2018) with an added contact round.

Stage 1: Individual production. Subjects performed a real-effort task, stuffing envelopes, individually for 12 minutes under a piece rate. Individual productivity was recorded and is the productivity measure shown to decision makers in Stage 3.

Stage 2: Contact (randomized). Subjects were randomly paired within session and performed the same task jointly for 12 minutes, paid collectively through a joint production function. Pairs were encouraged to interact. The treatment variable Contact_i equals 1 if subject i's partner was an international student. The pairing satisfies the classic conditions for effective contact (cooperation, equal status, and a common goal; Allport 1954) but is brief: a single twelve-minute episode, with most conversation concentrated in the first minute (introductions and a brief discussion of how to proceed) and at the end.

Stage 3: Partner choice (outcome). Subjects were told they *could* be invited to a future session. We used the conditional deliberately to avoid deception, since no future session actually took place. Subjects were asked to choose their preferred partner in 30 binary choices. For each candidate, three pieces of information were displayed: name (signaling nationality; foreign names are rendered in katakana), Stage-1 productivity, and an available time slot (included to reduce the salience of nationality). Of the 30 choices, 23 paired a Japanese with a foreign candidate and constitute the discrimination measure; the remaining 7 (4 Japanese–Japanese, 3 foreign–foreign) are decoys. Productivity differences between candidates span -20 to $+20$ envelopes. Foreign names were Chinese, Indian, and Korean, representing the most frequent nationalities within the student population. Choices were incentivized: participants were told that one of their decisions might be selected for implementation and that they might be invited back to carry it out in a future session. No participant was ultimately invited back.

Subjects were paid a show-up fee of JPY 1,000 (≈ USD 6.20). Bonus payment was determined by randomly selecting one of four components (individual production, joint production, dictator-game sender, dictator-game receiver).

2.3 Pre-registration

Hypotheses, outcome definitions, estimating equations, and a power analysis were pre-registered at the AEA RCT Registry on 13 March 2026, before data collection. The pre-analysis plan specified two primary hypotheses: contact reduces discrimination (H1) and contact reduces the willingness to pay for same-ethnicity partners (H2); and one secondary hypothesis: demand for discrimination is price-elastic, replicating Hedegaard and Tyran (2018) (H3).

3 Empirical strategy

Our primary estimator exploits the full choice panel (23 Japanese–foreign choices per subject; 164 × 23 = 3,772 observations). Let $F_{ij} = 1$ if subject i chooses the foreign candidate in choice situation j, and let $\Delta Prod_j$ denote the productivity difference (foreign minus Japanese). We estimate the logit demand curve

$$\Pr(F_{ij} = 1) = \Lambda(\alpha + \beta\,\Delta Prod_j + \gamma\,\text{Contact}_i + \delta\,(\Delta Prod_j \times \text{Contact}_i))\,, \tag{1}$$

with standard errors clustered at the participant level. β captures price sensitivity; γ and δ capture the contact effect on the level and slope of the demand curve. The willingness to pay for a Japanese partner is the productivity difference at which a subject is indifferent:

$$\text{WTP}_{\text{no contact}} = -\frac{\alpha}{\beta}, \qquad \text{WTP}_{\text{contact}} = -\frac{\alpha+\gamma}{\beta+\delta}, \qquad \Delta\text{WTP} = \text{WTP}_{\text{contact}} - \text{WTP}_{\text{no contact}}. \tag{2}$$

For ΔWTP we report bootstrap standard errors (resampling participants, stratified by arm). As pre-specified, we also estimate the same specification as a linear probability model as robustness check, and a participant-level regression of the share of discriminatory

choices on treatment.

4 Results

4.1 Discrimination is widespread and price-sensitive

Table A.2 reports balance on pre-treatment covariates between arms. The arms are well balanced on gender, age, and field of study; two education/field cells (holding a Master's degree, studying Law) differ at the 10% level, as expected by chance across many cells, and the design is randomized.

Table 1 reports the simplest measure: across the 23 Japanese–foreign choices, the share of subjects who chose the foreign candidate in fewer than half. The 23 choices span the productivity differences $-20, -10, -9, \ldots, -1, 0, +1, \ldots, +9, +10, +20$ (foreign minus Japanese): 22 form symmetric $\pm$ pairs and one is at equal productivity. In the control group this share is 45.4% (SE = 4.8 pp). About one half of subjects avoid the foreign partner in most choice situations, even when doing so is costly.

Table 1: Share of foreign partner

	Share of foreign < 0.5
	(1)
Treatment	-0.025
	(0.082)
Constant	0.454***
	(0.048)
No. obs.	164
RMSE	0.50

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 2 reports the logit demand curve (Equation 1). Price sensitivity is strong and precisely estimated: $\hat{\beta} = 0.150$ (SE = 0.016). In the linear probability model (Appendix Table A.3), a one-envelope increase in the foreign candidate's productivity advantage raises the probability of choosing them by 2.8 percentage points (SE = 0.2 pp). Discrimi-

nation is far from lexicographic: subjects trade off their preference for a Japanese partner against its price, which replicates the central finding of Hedegaard and Tyran (2018) (H3).

Table 2: Probability of choosing foreign partner (logit)

	Pr(Foreign = 1) (1)
$\Delta Prod$ (β)	0.150***
	(0.016)
Treatment (γ)	0.110
	(0.163)
$\Delta Prod\times$ Treatment (δ)	-0.014
	(0.027)
Constant (α)	-0.358***
	(0.098)
No. obs.	3,772
RMSE	0.44

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. Greek letters in parentheses represent the corresponding coefficients from Equation 1.

The constant in column 1 implies that at equal productivity ($\Delta Prod = 0$), a control-group subject chooses the foreign candidate with probability $\Lambda(-0.358) = 41.1\%$. Converting the demand curve into willingness to pay via Equation (2), the average control-group subject requires the foreign candidate to be 2.39 envelopes more productive (SE = 0.72) before being indifferent (Table 3). Relative to the average displayed candidate productivity of 40 envelopes, this is $2.39/40 \approx 6\%$, comparable to the 8% of income measured by Hedegaard and Tyran (2018) in Denmark.

4.2 The effect of contact is imprecisely estimated

The contact coefficients in Table 2 are individually and jointly insignificant: $\hat{\gamma} = 0.110$ (SE = 0.163) and $\hat{\delta} = -0.014$ (SE = 0.027), and a Wald test of the pre-registered joint restriction $\gamma = \delta = 0$ does not reject ($\chi^2_2 = 0.75$, $p = 0.69$; Table A.4). Table 3 translates the coefficients into the pre-registered headline quantity: willingness to pay is 2.39 (SE = 0.72) without contact and 1.82 (SE = 1.01) with contact. The difference, ΔWTP $= -0.57$

(SE = 1.22), is a 24% reduction relative to the control mean, in the direction predicted by contact theory, but statistically indistinguishable from zero ($p \approx 0.65$).

Table 3: Willingness to pay

	Mean	SD
$\text{WTP}_{\text{no contact}}$	2.39***	(0.717)
$\text{WTP}_{\text{contact}}$	1.82*	(1.006)
ΔWTP	-0.569	(1.223)

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. Willingness-to-pay is defined as the difference in productivity (counted as a number of envelopes) necessary to make subjects indifferent between the Japanese and foreigner candidates. The average displayed candidate productivity is 40 envelopes.

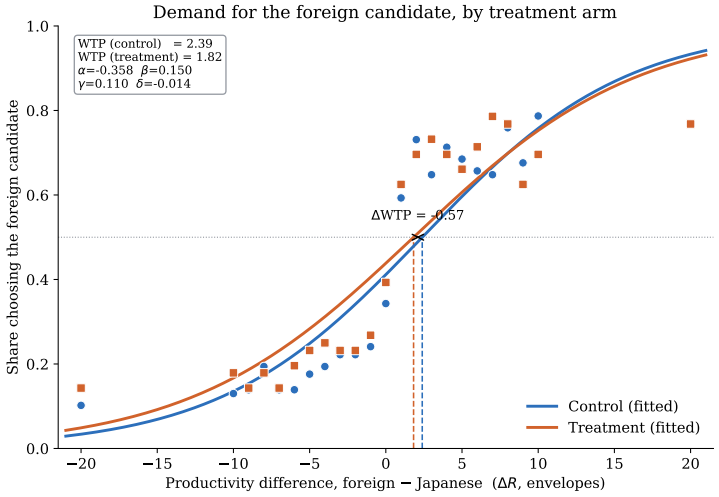

Figure 1: Demand curves and implied willingness to pay by treatment group

Notes: Share and fitted probability of choosing the foreign candidate as a function of the productivity difference (foreign minus Japanese), by treatment group. Vertical dashed lines mark the implied indifference points (WTP); the horizontal gap between them is $\Delta\text{WTP} \approx -0.57$.

Two features of this estimate matter for its interpretation. First, the confidence interval is wide: $[-2.94, +1.86]$ in productivity units, from a reduction larger than the control mean itself to a moderate increase. We can rule out effects as large as our pre-registered minimum detectable effect, a reduction of roughly 3.5 productivity units calibrated under an assumed baseline WTP of 10. But the realized baseline (2.39) is far lower, so effects of plausible relative magnitude, such as the 24% point estimate or effect sizes near the

meta-analytic average for contact interventions (Cohen's $d = 0.29$ in Clochard 2026), lie well inside the interval. Detecting an effect of the size we observe at conventional levels would require the standard error of ΔWTP to be roughly 4 times smaller, and hence a sample more than ten times larger. Second, the intervention was minimal by design: one twelve-minute episode of cooperative production. The estimate speaks to a *light-touch* contact treatment in this setting, not to contact in general. The data are consistent with contact reducing the price of prejudice, with no effect, and with a small increase.

4.3 Heterogeneity by outgroup identity

Table 4 disaggregates the choice situations by the origin of the foreign candidate's name (Chinese, Indian, Korean). Baseline discrimination differs markedly: at equal productivity, the constant is -0.585 (SE = 0.082) for Indian names but -0.257 (SE = 0.078) and -0.232 (SE = 0.082) for Chinese and Korean names, respectively. Discrimination is strongest against Indian-named candidates. Price sensitivity is similar across origins, and the contact coefficients remain insignificant in every subsample.

Table 4: Discrimination by origin of the foreign candidate (logit)

	Pr(Foreign = 1)			
	Baseline	Chinese	Indian	Korean
	(1)	(2)	(3)	(4)
$\Delta Prod$ (β)	0.150***	0.182***	0.179***	0.122***
	(0.016)	(0.019)	(0.015)	(0.012)
Treatment (γ)	0.110	0.062	0.163	0.122
	(0.163)	(0.131)	(0.136)	(0.141)
$\Delta Prod\times$ Treatment (δ)	-0.014	-0.031	-0.026	0.001
	(0.027)	(0.031)	(0.025)	(0.021)
Constant (α)	-0.358***	-0.257**	-0.585***	-0.232**
	(0.098)	(0.078)	(0.082)	(0.082)
No. obs.	3,772	1,312	1,312	1,148
RMSE	0.44	0.42	0.45	0.43

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In all columns, the outcome variable is a dummy variable for whether the subject chose the foreign candidate. Column 2 restricts the of choices to Japanese–Chinese pairs, column 3 to Japanese–Indian pairs, and column 4 to Japanese–Korean pairs. Greek letters in parentheses represent the corresponding coefficients from Equation 1.

Table 5 asks whether contact operates through the specific nationality encountered. Each column compares participants paired with a partner of a given nationality (treatment) against participants paired with a Japanese partner (control), restricting the choice situations to foreign candidates whose name signals that same nationality: column (2) is contact with a Chinese partner over Chinese-named candidates, and column (3) is contact with an Indian partner over Indian-named candidates. Point estimates are positive for Chinese contact (0.173, SE = 0.194) and negative for Indian contact (-0.197, SE = 0.255), but neither is significant and the subsamples are small. One candidate interpretation, suggested by the distribution of Japanese proficiency (Figure A.1), is language: Chinese students at The University of Osaka are recruited under language requirements and are close to fluent, while Indian students are not, so the treatment may partly correct a belief that foreigners cannot communicate in Japanese. Consistent with this, the interaction between treatment and a fluent partner's proficiency is the largest heterogeneous effect we find (0.369, SE = 0.230; Table A.5), though proficiency is almost perfectly collinear with nationality, so the two channels cannot be distinguished in this sample.

4.4 Exploratory: discrimination against women

The choice sets also generate within-gender and cross-gender candidate pairs, allowing an analysis we did not pre-register and present as exploratory. Table 6 estimates Equation 1 with the outcome "chose the female candidate." At equal productivity, the pooled constant of -0.467 (SE = 0.031) implies a 38.5% probability of choosing the female candidate, a larger deviation from indifference than the corresponding 41.1% for foreign candidates. The constants for female deciders (-0.411, SE = 0.053) and male deciders (-0.504, SE = 0.037) are similar: women avoid female partners almost as much as men do. The gender analysis uses the 20 choices per decider that pit a female against a male candidate, and—unlike the nationality analysis—pools Japanese deciders with all non-Japanese deciders (counting only each foreign participant's first session, since most attended more than once). This yields 187 deciders and $187 \times 20 = 3{,}740$ choices in the pooled column, $69 \times 20 = 1{,}380$ for female and $118 \times 20 = 2{,}360$ for male deciders.

Table 5: Treatment effect for specific nationality (logit)

	Pr(Foreign = 1)		
	Baseline	Chinese contact	Indian contact
	(1)	(2)	(3)
$\Delta Prod$	0.150***	0.182***	0.179***
	(0.016)	(0.026)	(0.020)
Treatment	0.110	0.173	-0.197
	(0.163)	(0.194)	(0.255)
$\Delta Prod\times$ Treatment	-0.014	-0.036	0.074
	(0.027)	(0.043)	(0.054)
Constant	-0.358***	-0.257*	-0.585***
	(0.098)	(0.107)	(0.113)
No. obs.	3,772	1,152	1,008
RMSE	0.44	0.42	0.44

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In all columns, the outcome variable is a dummy variable for whether the subject chose the foreign candidate. Column 2 restricts the of choices to Japanese–Chinese candidate pairs and to subjects whose partner was either Japanese or Chinese, column 3 to Japanese–Indian candidate pairs and to subjects whose partner was either Japanese or Indian. Greek letters in parentheses represent the corresponding coefficients from Equation 1.

Table 6: Contact with women (logit)

	Pr(Female = 1)		
	Pooled	Female	Male
	(1)	(2)	(3)
$\Delta Prod$ (β)	0.179***	0.179***	0.180***
	(0.007)	(0.009)	(0.009)
Treatment (female contact) (γ)	-0.040	-0.035	-0.031
	(0.058)	(0.112)	(0.066)
$\Delta Prod\times$ Treatment (female contact) (δ)	-0.007	0.003	-0.012
	(0.011)	(0.016)	(0.014)
Constant (α)	-0.467***	-0.411***	-0.504***
	(0.031)	(0.053)	(0.037)
No. obs.	3,740	1,380	2,360
RMSE	0.44	0.43	0.44

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In all columns, the outcome variable is a dummy variable for whether the subject chose the female candidate. In column 2, the sample is restricted to female subjects, in column 3 to male subjects. Greek letters in parentheses represent the corresponding coefficients from Equation 1.

Table 7 contains a cautionary result. Pooling all subjects, being paired with an opposite-gender partner in the contact round appears to *increase* the probability of choosing opposite-gender candidates (0.207, SE = 0.078, significant at 5%). But the effect vanishes within gender (column 2: −0.035; column 3: −0.031): the pooled coefficient is an artefact of unequal gender composition across pairing types (Table A.6), not a treatment effect. Because gender of the contact partner was not the randomized dimension of the design, we flag this exploratory analysis as descriptive.

Table 7: Contact with other gender (logit)

	Pr(Other gender = 1)		
	Pooled	Female	Male
	(1)	(2)	(3)
$\Delta Prod$ (β)	0.172***	0.183***	0.180***
	(0.007)	(0.014)	(0.009)
Treatment (other gender) (γ)	0.207**	-0.035	-0.031
	(0.078)	(0.112)	(0.066)
$\Delta Prod\times$ Treatment (other gender) (δ)	-0.010	-0.003	-0.012
	(0.009)	(0.016)	(0.014)
Constant (α)	-0.273***	0.446***	-0.504***
	(0.052)	(0.098)	(0.037)
No. obs.	3,740	1,380	2,360
RMSE	0.45	0.43	0.44

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. In all columns, the outcome variable is a dummy variable for whether the subject chose the candidate of the opposite gender as themselves. In column 2, the sample is restricted to female subjects, in column 3 to male subjects. Greek letters in parentheses represent the corresponding coefficients from Equation 1.

4.5 Secondary outcomes

Dictator game. Out of a JPY 1,000 endowment, treated subjects kept JPY 575 on average against JPY 612 in control, a difference of −36.7 (SE = 34.5) — more generous after contact, but insignificant (Table A.7). This comparison is hard to interpret cleanly, because contact changes both the treatment (paired with a foreign rather than a Japanese partner) and the identity of the recipient, so it does not isolate a contact effect on generosity toward a fixed target.

Charitable giving. In a post-task choice, participants selected which of two NGOs should receive a JPY 10,000 donation if they were randomly drawn at the end of the experiment. The options were the Japan Association for Nature Conservation (a neutral alternative, described as protecting endangered species and Japan's natural environment) and the Japan Student Services Organization (JASSO, described as a body that provides scholarships and support for international students and promotes international exchange). The pre-registered outcome is the share choosing JASSO by treatment arm. Contact does not move it: 63.9% of control subjects and 62.5% of treated subjects chose JASSO (χ^2 test $p = 0.86$; Table A.8).

Discrimination shares. As pre-registered, we also summarize discrimination at the participant level (Table 8). For each subject we compute the share of choices in which they "pay" for a Japanese partner—declining a strictly more productive foreign candidate ($\Delta Prod > 0$)—and, symmetrically, the share in which they decline a strictly more productive Japanese candidate ($\Delta Prod < 0$); the net share is the difference. In the control group the net share is 0.130 (SE = 0.038), confirming systematic anti-foreign discrimination. The contact effect is insignificant on the net share (-0.036, SE = 0.064), the against-foreign share (-0.009, SE = 0.052), and the against-Japanese share (0.027, SE = 0.040), consistent with the choice-panel estimates.

Immigration attitudes. The pre-registered post-task survey included an immigration-attitude item and a generalized-trust item.[3] Contact does not significantly move either (Table A.9): the treatment effect is 0.111 (SE = 0.145) on the 1–5 immigration item and 0.603 (SE = 0.363) on the 0–10 trust item. Because neither the attitude first stage nor the discrimination reduced form is distinguishable from zero, the pre-registered mediation analysis (contact $\rightarrow$ attitudes $\rightarrow$ discrimination) is uninformative in this sample and we do not report it.

[3] "Do you agree or disagree that immigration to Japan should be increased?" (1 = strongly disagree to 5 = strongly agree) and "To what extent do you trust people that you do not know?" (0 = do not trust at all to 10 = trust completely). Higher values indicate more favorable attitudes to immigration and higher trust.

Table 8: Effect of contact on the shares of discriminatory choices

	Discrimination shares		
	Net (Foreign - Japanese)	Against Foreign	Against Japanese
	(1)	(2)	(3)
Treatment	-0.036	-0.009	0.027
	(0.064)	(0.052)	(0.040)
Constant	0.130***	0.303***	0.173***
	(0.038)	(0.030)	(0.023)
No. Obs.	164	164	164
RMSE	0.39	0.31	0.24

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome variable in column 2 is the share of choices when subjects chose the Japanese candidate, conditional on the foreign candidate being more productive, measured at the subject level. The outcome variable in column 3 is the similar share, but reversed (choice of foreign candidate when Japanese candidate was more productive). The outcome variable in column 1 is the difference between the two outcomes.

Heterogeneity. We pre-registered a generic machine-learning analysis of treatment-effect heterogeneity (Chernozhukov et al., 2025). We find no evidence that the contact effect varies across participants, and a complementary analysis shows that the level of discrimination, though highly variable across individuals, is not predictable from the characteristics we observe. Appendix C reports both. One pre-registered item remains unaddressed: we did not collect a measure of prior contact with foreigners, so prior contact cannot be used as a moderator.

Experimenter Demand. A common concern in experiments on discrimination is that participants may infer the purpose of the study and adjust their behavior accordingly, generating experimenter-demand effects (Zizzo, 2010). To assess this possibility, immediately after the discrimination measure (and before the explicit survey questions), we asked participants an open-ended question about what they believed the purpose of the experiment was. We then manually and algorithmically classified their responses using the awareness measures described in Table A.10. As reported in Table A.11, very few participants correctly inferred the study's objective: only about 14% gave responses indicating awareness that the experiment concerned a nationality–productivity trade-off.

Moreover, none of the awareness measures differ significantly between the treatment and control groups, suggesting that the treatment did not make participants more likely to infer the purpose of the experiment.

5 Discussion

What we learn about measurement. The price-of-prejudice paradigm replicates. With 30 fully incentivized choices and no deception, we obtain a demand curve for discrimination that is precisely estimated, price-elastic, and quantitatively close to the original Danish results (WTP of roughly 6% of earnings versus 8%). The method identifies discrimination without incentivizing or priming it, and it does so in a context with a very different immigration history and a different outgroup. Interestingly, while Hedegaard and Tyran (2018) use a field experiment, we show that we can replicate this measure in a laboratory experiment. The measurement instrument is portable and can serve as an outcome in future intervention studies.

What we learn about contact. The contact estimate is directionally consistent with the contact hypothesis and economically meaningful at face value (a 24% reduction in WTP), but the design cannot distinguish it from zero. A credible test of contact in this paradigm requires some combination of (i) a substantially larger sample, on the order of ten times ours for an effect of the size we observe, (ii) a stronger or repeated contact intervention, or (iii) a setting with higher baseline discrimination, where the same relative effect is a larger absolute quantity. The realized baseline WTP (2.39 units) was well below the value assumed in our power calculations (10 units), which mechanically inflated the relative MDE; future designs should calibrate power to conservative baseline levels.

Taste-based or statistical? Because individual productivity is displayed, the residual preference for Japanese partners is naturally read as taste-based. It is, however, consistent with statistical discrimination over unobserved attributes of the pair interaction: beliefs that a Japanese partner will exert more effort in the joint task, for instance, or that

communication will be easier. The proficiency pattern in Section 4.3 suggests beliefs about language may matter. Our design does not separate these channels; a treatment that varies information about partner attributes other than productivity could.

The gender finding. The discrimination against female candidates, larger than against foreign candidates and shared by female deciders, was not predicted and should be treated as hypothesis-generating. It also reinforces the measurement point: the paradigm picks up discrimination on dimensions the experimenters did not emphasize, at magnitudes large enough to matter.

6 Conclusion

We replicate the price of prejudice in Japan: about half of subjects pay to avoid working with foreigners, the average price paid is about 6% of earnings, and demand for discrimination slopes downward. A brief cooperative contact intervention moves the point estimate of the willingness to pay in the predicted direction, but the effect is statistically insignificant and the design underpowered for plausible effect sizes. The validated measurement instrument, the estimated baseline, and the power benchmark are the inputs a better-powered or stronger-treatment experiment will need, and we see running one as the natural next step.

References

Achard, Pascal, Sabina Albrecht, Riccardo Ghidoni, Elena Cettolin and Sigrid Suetens. 2025. "Local exposure to refugees changed attitudes to ethnic minorities in the Netherlands." *The Economic Journal* 135(667):808–837.

Allport, Gordon W. 1954. *The Nature of Prejudice*. Cambridge, MA: Addison-Wesley.

Bajzíková, Stanislava and Lubomír Cingl. 2023. "Measuring stereotypes in effort tasks:

A multiple-price list approach." *Journal of Behavioral and Experimental Economics* 105:102032.

Becker, Gary S. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.

Bertrand, Marianne and Esther Duflo. 2017. Field experiments on discrimination. In *Handbook of economic field experiments*. Vol. 1 Elsevier pp. 309–393.

Chernozhukov, Victor, Mert Demirer, Esther Duflo and Iván Fernández-Val. 2025. "Fisher–Schultz Lecture: Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, With an Application to Immunization in India." *Econometrica* 93(4):1121–1164.

Clochard, Gwen-Jirō. 2025. "Using a brief contact to improve trust in the police by the youth." *Journal of Behavioral and Experimental Economics* 117:102376.

Clochard, Gwen-Jiro. 2026. Reducing Prejudice: A Meta-Analysis and Experiment on the Contact Hypothesis. Technical report Working paper.

Clochard, Gwen-Jirō, Guillaume Hollard and Omar Sene. 2026. "Bringing contact interventions to the lab: Effects of brief bilateral discussions on interethnic trust in Senegal." *World Development* 199:107247.

Durante, Federica and Susan T Fiske. 2017. "How social-class stereotypes maintain inequality." *Current opinion in psychology* 18:43–48.

Elwert, Felix, Tamás Keller and Andreas Kotsadam. 2023. "Rearranging the desk chairs: A large randomized field experiment on the effects of close contact on interethnic relations." *American Journal of Sociology* 128(6):1809–1840.

Finseraas, Henning and Andreas Kotsadam. 2017. "Does personal contact with ethnic minorities affect anti-immigrant sentiments? Evidence from a field experiment." *European Journal of Political Research* 56(3):703–722.

Finseraas, Henning, Torbjørn Hanson, Åshild A Johnsen, Andreas Kotsadam and Gaute Torsvik. 2019. "Trust, ethnic diversity, and personal contact: A field experiment." *Journal of Public Economics* 173:72–84.

Greiner, Ben. 2015. "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of the Economic Science Association* 1(1):114–125.

Hedegaard, Morten Størling and Jean-Robert Tyran. 2018. "The price of prejudice." *American Economic Journal: Applied Economics* 10(1):40–63.

Hjort, Jonas. 2014. "Ethnic divisions and production in firms." *The Quarterly Journal of Economics* 129(4):1899–1946.

Hjort, Jonas, Changcheng Song and Chris Yenkey. 2026. "Ethnic investing and the value of firms." *Management Science* 72:6419–6432.

Igarashi, Akira and Ryota Mugiyama. 2023. "Whose tastes matter? Discrimination against immigrants in the Japanese labour market." *Journal of Ethnic and Migration Studies* 49(13):3365–3388.

JOSHI, Ratala Dinesh Prasad and Satoko Tabata. 2021. "Barriers to education for migrant children in Japan at mainstream schools: From a rights-based perspective." *Journal of Human Security Studies* 10(2):109–129.

Lowe, Matt. 2025. "Has intergroup contact delivered?" *Annual Review of Economics* 17:321–344.

Paluck, Elizabeth Levy, Seth A Green and Donald P Green. 2019. "The contact hypothesis re-evaluated." *Behavioural Public Policy* 3(2):129–158.

Pettigrew, Thomas F and Linda R Tropp. 2006. "A meta-analytic test of intergroup contact theory." *Journal of Personality and Social Psychology* 90(5):751–783.

Scacco, Alexandra and Shana S Warren. 2018. "Can social contact reduce prejudice and discrimination? Evidence from a field experiment in Nigeria." *American Political Science Review* 112(3):654–677.

Sugasawa, Takeru and Kei Harano. 2023. "A field experiment on discrimination against foreigners in the rental housing market in Japan examining the 23 wards of Tokyo." *Journal of the Japanese and International Economies* 69:101273.

Wager, Stefan and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association* 113(523):1228–1242.

Zizzo, Daniel John. 2010. "Experimenter demand effects in economic experiments." *Experimental Economics* 13(1):75–98.

A Additional tables and figures

Table A.1: Partner nationality

Nationality	**N**
Japanese (control)	108
Chinese	36
Indian	18
Ugandan	2

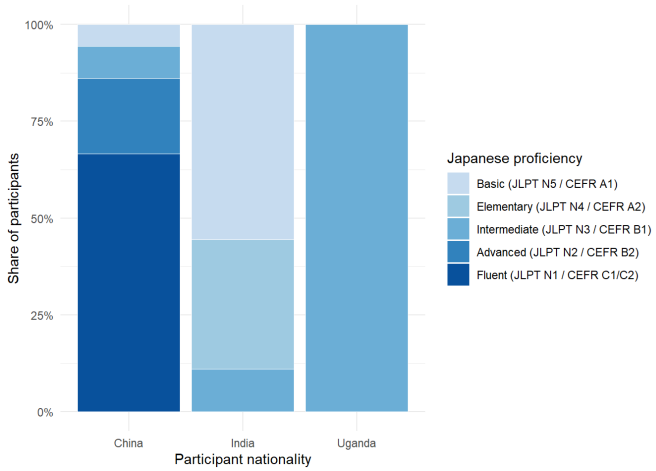

Figure A.1: Distribution of Japanese proficiency by nationality (non-Japanese participants)

Table A.2: Covariate balance between treatment and control groups

Covariate	Level	Control	Treatment	SMD
intro education	Master's degree	0.148	0.054	-0.318*
intro field of study	Law	0.046	0.000	-0.312*
intro field of study	Other	0.056	0.018	-0.201
participant gender	prefer not to say	0.019	0.000	-0.194
intro field of study	Humanities	0.019	0.054	0.189
intro field of study	Pharmacy	0.046	0.089	0.172
intro education	Completed primary education	0.009	0.000	-0.137
intro education	Bachelor's degree	0.324	0.375	0.107
participant gender	female	0.324	0.375	0.107
intro birth year		2002.509	2002.804	0.105
intro field of study	Medicine	0.093	0.125	0.104
intro field of study	Economics	0.056	0.036	-0.095
intro education	Completed secondary education	0.463	0.500	0.074
intro education	PhD or equivalent	0.009	0.018	0.074
participant gender	male	0.657	0.625	-0.068
intro field of study	Foreign languages	0.046	0.036	-0.053
intro field of study	Science	0.065	0.054	-0.048
intro education	Completed college of technology	0.028	0.036	0.045
intro field of study	Dentistry	0.065	0.071	0.026
intro field of study	Engineering	0.509	0.518	0.017
intro education	Prefer not to say	0.019	0.018	-0.005

Table A.3: Probability of choosing foreign partner (OLS)

	Foreign (1)
$\Delta Prod$	0.028***
	(0.002)
Treatment	0.020
	(0.032)
$\Delta Prod\times$ Treatment	-0.001
	(0.003)
Constant	0.431***
	(0.019)
No. obs.	3,772
R^2	0.203

* p < 0.10, ** p < 0.05, *** p < 0.01. The outcome variable is a dummy variable for whether the subject chose the foreign candidate.

Table A.4: Wald test of the joint contact effect ($\gamma = \delta = 0$) in the logit demand curve

Res. Df	Df	χ^2	Pr($> \chi^2$)
3768			
3770	-2	0.750	0.687

Notes: Joint test that the contact level (γ) and contact $\times$ productivity-difference slope (δ) coefficients in Table 2 are both zero, with standard errors clustered by participant.

Table A.5: Treatment effect by actual foreign partner's Japanese proficiency (logit)

	Pr(Foreign = 1)
$\Delta Prod$	0.150***
	(0.016)
$\Delta Prod\times$ Contact	−0.014
	(0.027)
Contact × Elementary	0.460
	(0.443)
Contact × Intermediate	0.061
	(0.355)
Contact × Advanced	−0.194
	(0.352)
Contact × Fluent	0.369
	(0.230)
Constant	−0.387***
	(0.091)
No. obs.	3,772
RMSE	0.43

Notes: Logit of choosing the foreign candidate. Treatment (contact with a foreign partner) is interacted with the partner's self-reported Japanese proficiency; the omitted reference is a Basic-proficiency partner (and all control partners are coded Basic). Standard errors clustered by participant. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Treatment with gender

Subject Gender	**Partner Gender**	**N**
Female	Male	47
Male	Male	71
Female	Female	22
Male	Female	47

Table A.7: Dictator game

	Amount kept	
	Mean	SD
Control	611.67	(205.08)
Treatment	574.98	(211.67)
Difference	-36.69	(34.48)

Notes: The variable represents the amount kept in a Dictator game played with the partner from the contact phase.

Table A.8: Charitable giving: Choice of NGO supporting international students.

Share (Control)	Share (Treatment)	p-value
0.639	0.625	0.861

Notes: The variable displayed is the share of subjects in each treatment who chose the Japan Student Services Association (JASSO), an organization that provides scholarships and support for international students, relative to the Japan Association for Nature Conservation, described as protecting endangered species in Japan.

Table A.9: Post-task attitudes by treatment group

	Control mean	Treatment effect	No. obs.
Immigration attitude (1–5)	2.389	0.111	164
		(0.145)	
Trust in strangers (0–10)	3.880	0.603	164
		(0.363)	

Notes: OLS of each post-task survey response on the contact treatment indicator; heteroskedasticity-robust (HC1) standard errors in parentheses. "Control mean" is the intercept. Higher values follow the orientation of the post-task questionnaire. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Experimenter demand: measures definition

Measure	**Computation**	**Interpretation**
Manual: mentions nationality	Manual coding	Participant explicitly mentioned foreign / Japanese nationality.
Manual: mentions nationality and productivity	Manual coding	Participant explicitly mentioned both foreign / Japanese nationality and productivity.
Substantive response	Response has at least 10 non-whitespace characters.	
Partner-choice	Response contains at least one partner / person keyword and at least one choice / "work-with".	Participant understood that the experiment involved choosing or working with partners.
Nationality partner	Response contains a nationality / Japanese / foreign keyword and also indicates a partner-choice or work-with context.	Participant appeared aware that nationality or Japanese / foreign partner identity was part of the study.
Nationality–productivity tradeoff	Response satisfies the nationality / Japanese-partner awareness measure and also mentions productivity, ability, reward, price, or another tradeoff-related concept.	Participant appeared closest to the target objective: a nationality-related partner-choice tradeoff involving productivity or payoff.

Table A.11: Differences in Awareness between Treatment groups

Measure	**Control mean**	**Treatment mean**	**Difference (T − C)**	p
Manual: mentions nationality	0.148	0.107	−0.041	0.465
Manual: mentions nationality and productivity	0.074	0.107	0.033	0.472
Substantive response	1.000	1.000	0.000	1.000
Partner-choice	0.620	0.607	−0.013	0.869
Nationality-partner	0.250	0.286	0.036	0.622
Nationality–productivity tradeoff	0.130	0.161	0.031	0.587

B Experimental materials

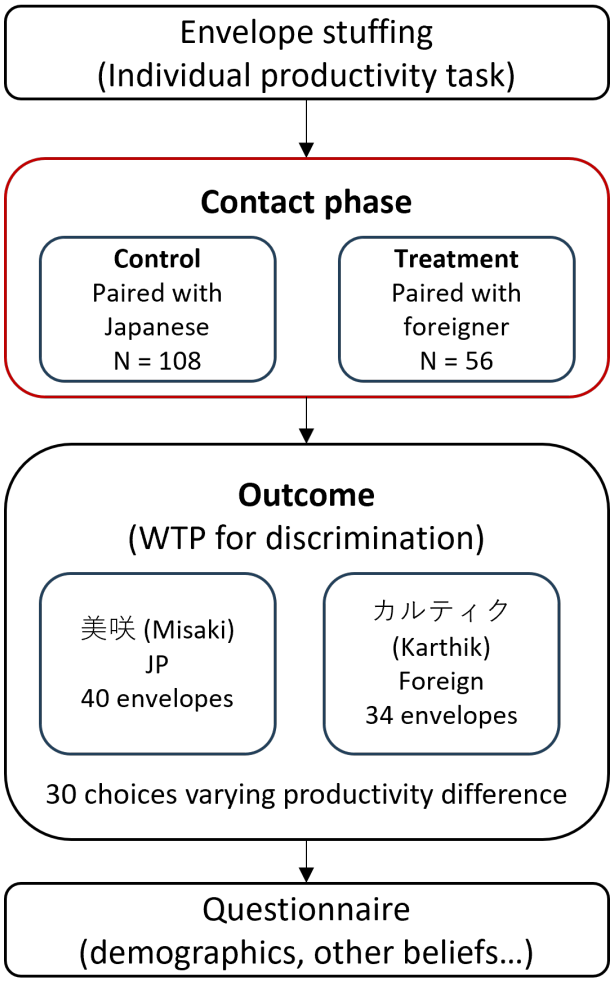

Figure B.2: Experimental flow

Figure B.3: Example of a partner choice

C Heterogeneity analyses

We report two distinct heterogeneity exercises. The first asks whether the *effect of contact* varies across participants. The second asks whether the *level* of discrimination, the price of prejudice itself, is predictable from observable characteristics, setting the treatment aside. The first is a causal question that exploits the randomization; the second is descriptive.

Heterogeneity in the contact effect. Following the pre-analysis plan, we apply the generic machine-learning procedure of Chernozhukov et al. (2025). We summarize each subject by a participant-level discrimination measure (the net share of choices in which they pay to avoid the foreign candidate, and, as an alternative, the share of foreign candidates chosen). We proxy the baseline and treatment-effect functions with machine learning fit on a random auxiliary half-sample and evaluated on the other half, using two learners (a random forest and an elastic net) and aggregating over 100 random splits with the median adjustment of Chernozhukov et al. (2025). The best linear predictor of the conditional average treatment effect recovers the average effect, but its heterogeneity loading is statistically indistinguishable from zero (median-adjusted p close to one for both learners), and the sorted group average treatment effects of the most and least affected groups cannot be told apart. A causal forest (Wager and Athey, 2018), estimated as an independent cross-check, agrees: its calibration test for differential prediction does not reject the null of no heterogeneity (p of about 0.7), and the implied average effect matches our main estimate. With 164 participants and an average effect that is itself imprecisely estimated, power to detect heterogeneity is low, so we read these results as no evidence of heterogeneity rather than evidence of its absence.

Heterogeneity in the price of prejudice. There is substantial variation across subjects in how much they discriminate. The participant-level net discrimination share has a standard deviation of about 0.39 around a mean near 0.13, so that some subjects almost always pay to avoid the foreign candidate while others never do. This variation is, however, essentially unrelated to the characteristics we observe. We predict the individual

discrimination measure from pre-treatment covariates (gender, birth year, education, field of study, and the subject's own Stage-1 productivity), using both a linear best predictor and a random forest, in the control group alone and in the full sample. The cross-validated out-of-sample R^2 is approximately zero in every case, and the random forest predicts no better than the sample mean. No single covariate, including own productivity, robustly predicts discrimination once we account for overfitting and for near-empty covariate cells. We conclude that the price of prejudice is highly heterogeneous across individuals but idiosyncratic with respect to the observables available to us.