Supporting Information

"Do Survey Experiments Capture Real-World Behavior? External Validation
of Conjoint and Vignette Analyses with a Natural Experiment"

Jens Hainmueller[*], Dominik Hangartner[†][‡], Teppei Yamamoto[§]

[*]*Department of Political Science, Stanford University*
[†]*Department of Methodology, London School of Economics*
[‡]*Institute of Political Science, University of Zurich*
[§]*Department of Political Science, Massachusetts Institute of Technology*

Contents

S0   Introduction

This Supporting Information is structured as follows: In the first section we provide more background information about the naturalization referendums. The second section presents evidence suggesting that immigration-related preferences remained fairly stable from the time when the use of naturalization referendums ended and the time when we fielded our survey. The third section provides details about the survey sample. The fourth section provides details about the experimental design. The fifth section reports additional results and robustness checks for the main analysis. The last section reports additional results about the survey engagement in the different experimental designs.

S1   Behavioral Benchmark: Naturalization Referendums

In Switzerland, each municipality autonomously decides on the naturalization applications of its foreign residents who seek Swiss citizenship (for more details on the Swiss naturalization procedure, see [1]). We focus on the group of municipalities that until 2003 used referendums with closed ballots to decide on naturalization requests. A typical naturalization referendum involved two stages. Local voters first received in the mail the ballot and an official voting leaflet that explained the pending naturalization request with a detailed description of each immigrant applicant including information about his or her age, gender, education, origin, language skills and integration status. Figure S1 shows an anonymized example of a typical voting leaflet. Figure S2 provides an English translation. Voters then cast a secret ballot on each individual request and applicants with a majority of "yes" votes were granted Swiss citizenship.

Figure S1: Sample leaflet sent out to voters (names blacked out)

Traktandum 11

Einbürgerungsgesuch der ██████████ **, 1986, ledig, italienische Staatsangehörige,** ████████████ **6430 Schwyz**

A. Bericht

Mit Datum vom 14. Februar 2002 stellt die italienische Staatsangehörige, ██████████ █████, geboren am 10. September 1986 in Schwyz, wohnhaft in Schwyz, ██████████ ██, das Gesuch um Erteilung des Bürgerrechtes der Gemeinde Schwyz.

Die Gesuchstellerin ist am 10. September 1986 in Schwyz geboren als Tochter des ███ ██████████ und der ████████████ Sie ist seit Geburt in Schwyz angemeldet und wohnt mit ihren Eltern zusammen in Schwyz. ████████████ besitzt einen unbescholtenen Leumund. Sie hat den Kindergarten, die obligatorische Primar- und ein Jahr Realschule in Schwyz besucht. Momentan besucht sie die dritte Sekundarklasse im Schulhaus Rubiswil in Ibach. In ihrer Freizeit verdient sich als Verkäuferin im Warenhaus Manor ihr eigenes Taschengeld. Nach Abschluss der Schule beabsichtigt ██████████ ██, eine kaufmännische Lehre in einem Reisebüro zu absolvieren.

Der Vater der Gesuchstellerin kam 1966 und die Mutter 1971 in die Schweiz. ████ ██ ████ ███████████ ████████████. Die Gesuchstellerin ist zusammen mit vier Geschwistern aufgewachsen. Ihr Heimatland Italien kennt sie nur von den Ferien mit ihren Eltern. Sie könnte sich nicht vorstellen, nach Italien zu ziehen und dort zu wohnen. Die Gesuchstellerin spricht unseren Dialekt absolut akzentfrei. Sie ist mit unseren Sitten und Gebräuchen vertraut und kann als völlig integriert betrachtet werden. Die gemachten Erhebungen brachten nichts zum Vorschein, was gegen eine Aufnahme der Bewerberin ins Schweizer Bürgerrecht sprechen würde. Die formellen und materiellen Voraussetzungen für die Einbürgerung sind erfüllt.

Das Bundesamt für Polizeiwesen hat der Gesuchstellerin am 28. Januar 2002 die Einbürgerungsbewilligung ausgestellt. Diese ist drei Jahre gültig. Das Departement des Innern des Kantons Schwyz hat gegen die Einbürgerung nichts einzuwenden. Der Gemeinderat empfiehlt dem Bürgerrechtsgesuch zuzustimmen.
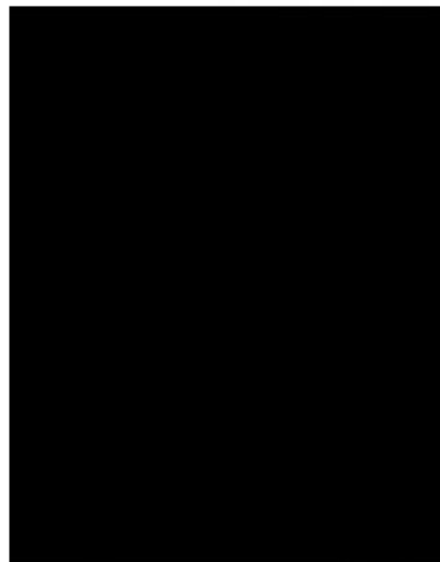
Figure S2: Sample leaflet sent out to voters (English translation)

Agenda item 11

Naturalization application of Applicant, **1986, unmarried, Italian citizen,** Street, **6430 Schwyz**

A. Report

On 14 February 2002, Applicant, Italian citizen, born on September 10, 1986 in Schwyz, domiciled in Schwyz, Street, applied for citizenship in the municipality of Schwyz.

The applicant was born on September 10, 1986 in Schwyz as the daughter of Applicant's Father and Applicant's Mother. She has been registered in Schwyz since birth and lives with her parents in Schwyz. Applicant has a good reputation. She attended kindergarten, primary school and one year of lower secondary school (*Realschule*) in Schwyz. Currently she attends the third year of upper secondary school (*Sekundarschule*) at 'Rubinswil' in Ibach. In her spare time, she earns pocket money as a shop assistant at the department store 'Manor'. After completing school, Applicant intends to take up an apprenticeship at a travel agency.

The applicant's father migrated to Switzerland in 1966, her mother in 1971. Applicant Redacted Information. The applicant grew up with 4 siblings. She knows Italy, her country of origin, only from vacations with her parents. She cannot imagine to move to Italy to live there. The applicant speaks our dialect with absolutely no accent. She is familiar with our culture and customs and can be considered fully integrated. Our enquiries did not reveal anything that would speak against granting Applicant Swiss citizenship. The substantial and formal requirements for naturalization are fulfilled.

The Federal Office of Police issued the applicant's naturalization permit on January 28, 2002. The permit is valid for three years. The Department of Home Affairs of the canton of Schwyz has no objection to the naturalization. The municipal council recommends approving the naturalization request.

We use a subset of the data compiled by Hainmueller and Hangartner ([1]) that contains applicant characteristics and voting outcomes for 1,503 recorded naturalization referendums held between 1970 and 2003 in the 44 Swiss municipalities that used secret ballot referendums with voting leaflets.[1] The majority of the data consists of naturalization referendums held

---

[1]The 44 municipalities are: Altdorf, Altendorf, Arth, Beckenried, Bühler, Buochs, Chur, Dallenwil, Davos, Einsiedeln, Emmen, Ennetmoos, Feusisberg, Freienbach, Gais, Galgenen, Gersau, Heiden, Hergiswil, Ingenbohl,

between 2000 and 2003. The behavioral data is recoded to match the survey attributes discussed below. We use these data to examine how applicant characteristics affect the outcome of naturalization referendums and thereby form the behavioral benchmark that we try to replicate with different survey experimental designs.

S2 Stability of Immigration Preferences

The use of naturalization referendums ended in 2003, whereas our survey was administered in 2014. We use two different data sets to examine if voters' preferences regarding immigration might have changed between these years.

First, we use annual panel data from the Swiss Household Panel (SHP)[2], to track changes in attitudes towards immigrants. The only immigration-related question in the SHP reads as follows: "Are you in favour of Switzerland offering foreigners the same opportunities as those offered to Swiss citizens, or in favour of Switzerland offering Swiss citizens better opportunities?". Answers were recorded on a three point scale as (1) foreigners and Swiss citizens should be offered *equal opportunities*, (0) *neither* or (-1) Swiss citizens should be offered *better opportunities*. We use the subset of, on average, $N = 1,395$ respondents per wave that reside in cantons that contain at least one target municipality. Figure S3 presents the SHP results. The trends over the years 1999 – 2009[3] are remarkably stable.

Küssnacht, Lachen, Malters, Morschach, Oberiberg, Reichenburg, Rothenthurm, Schübelbach, Schwyz, Speicher, St. Margrethen, Stans, Stansstad, Steinen, Teufen, Trogen, Tuggen, Unteriberg, Urnäsch, Walzenhausen, Wangen, Weggis, Wolfenschiessen, and Wollerau.

[2]The data is hosted at `http://forscenter.ch/fr/our-surveys/swiss-household-panel/`.

[3]Unfortunately, the question about opportunities for Swiss natives and foreigners was discontinued in 2010.

Figure S3: Stability of attitudes towards immigrants over time; Swiss Household Panel

Are you in favour of Switzerland offering foreigners the same opportunities as those offered to Swiss citizens, or in favour of Switzerland offering Swiss citizens better opportunities?
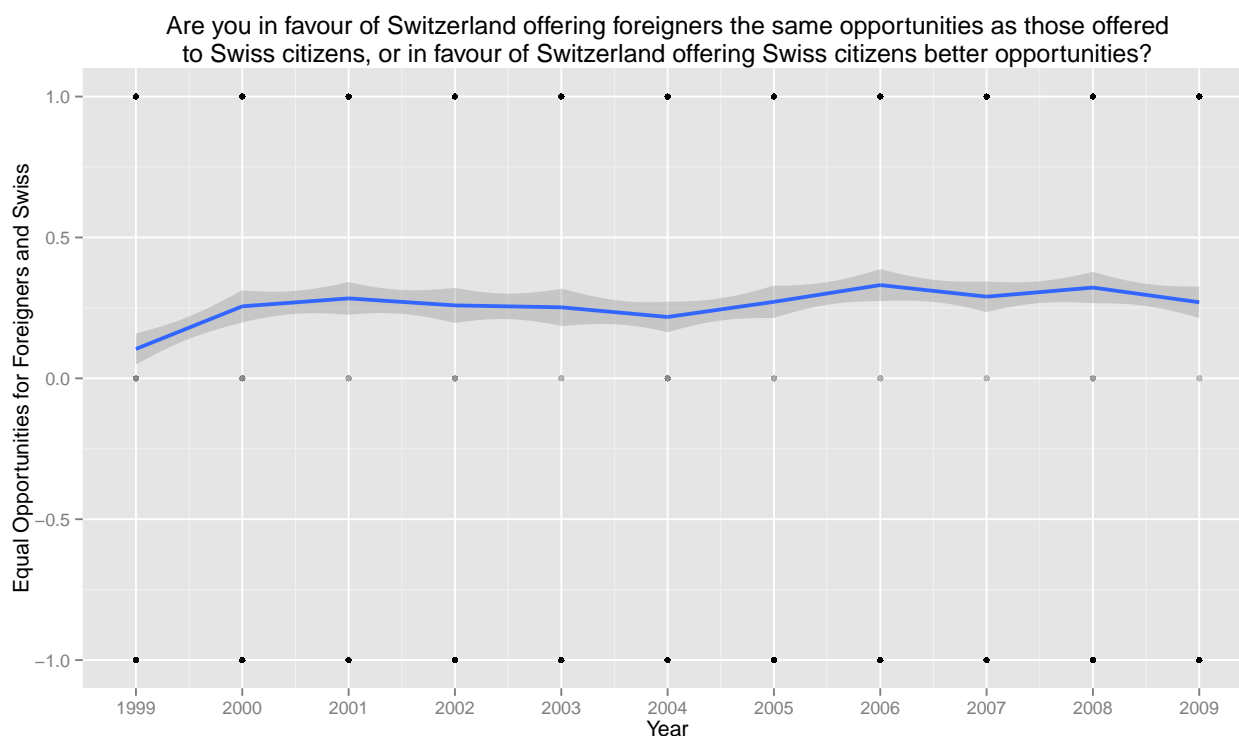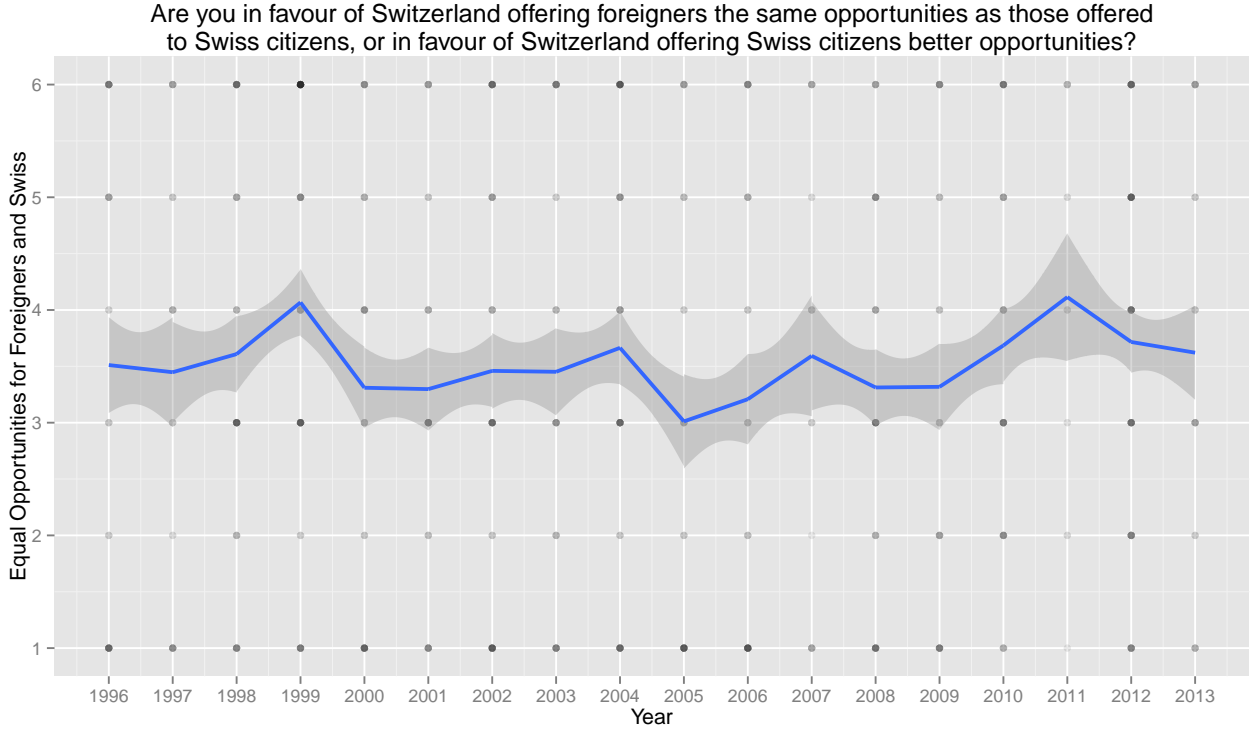
Figure shows year-to-year averages and corresponding 95 % confidence intervals. The variable measures attitudes towards immigrants over the years 1999–2009. Data: Swiss Household Panel, focusing on respondents from cantons that contain at least one target municipality. The sample size consists of, on average, $N = 1395$ respondents per year.

Second, we use the VOX survey[4], a cross-sectional post-referendum survey conducted about $3 - 6$ times per year with a sample size of approximately 1,000 respondents per wave. The only immigration-related question that is repeatedly asked in the VOX survey is identical to the one from the SHP but coded slightly differently insofar as answers were recorded on a six point scale from (1) Swiss citizens should be offered *better opportunities*, to (6) foreigners and Swiss citizens should be offered *equal opportunities*. We use the subset of, on average, $N = 104$ respondents per year that reside in one of the 44 target municipalities. Figure S4 presents the VOX results. While there is some year-to-year variance due to the small sample size, the overall trends over the years $1996 - 2013$ are remarkably stable.

---

[4]The data is hosted at `http://forsdata.unil.ch/projects/voxit/sondages.asp?`.

Figure S4: Stability of attitudes towards immigrants over time; VOX survey

Are you in favour of Switzerland offering foreigners the same opportunities as those offered
to Swiss citizens, or in favour of Switzerland offering Swiss citizens better opportunities?

Figure shows year-to-year averages and corresponding 95 % confidence intervals. The variable measures attitudes towards immigrants over the years 1996–2013. Data: Swiss post-referendum survey VOX, focusing on respondents from target municipality. The sample size consists of, on average, $N = 104$ respondents per year.

In summary, both tests provide some suggestive evidence that attitudes towards immigrants have remained fairly stable over the ten years separating the behavioral data from our survey experiment.

S3   Survey Design and Sample

Recruitment and Response Rate

Our main experiment was embedded in a survey that we conducted with the Swiss research firm gfs.bern. The field work took place between March 5 and July 25, 2014. The sampling design was a stratified random sampling. The recruitment was done by gfs.bern who contacted a stratified (by age and gender) random sample of 12,236 individuals in the target municipalities by telephone to invite them to participate in our online survey and collect baseline demographics and respondents' email addresses. Of these, 2,517 respondents agreed to participate in our online survey and were invited by email. Of those that expressed their willingness to participate, $N = 1,979$ respondents completed the survey, yielding a retention rate of 78.6% from telephone interview to online survey.[5] Overall, this corresponds to a participation rate of 20.6 % and a cumulative response rate 3 (RR3) as defined by AAPOR of 12.8 %. Note that this RR3 is

---

[5]All respondents who initially agreed to participate in the online survey were reminded twice per email and a third time per telephone in the two four weeks following the initial email invitation.

substantially higher than that of comparable online surveys. For example, a typical recent study conducted via Knowledge Networks, widely regarded as one of the best probability based online panels in the U.S., yields an RR3 of 2.8 % [2].

Sample Descriptives

Table S1 shows the respondent characteristics for the unweighted survey sample, the Swiss post-referendum study VOX, and the reweighted survey sample. The VOX survey is the best available survey data on the Swiss voting population.

   We see that the raw characteristics in our survey sample are quite close to the VOX survey. To address the small remaining differences we use entropy balancing [3] to reweight the survey sample based on the margins for age, gender, political interest, hypothetical participation in referendums, education, and employment to the margins computed from the VOX data. To create the margins for the reweighting procedure, we only focus on the VOX respondents that live in one of the target municipalities between 2003–2013. After reweighting, the observable characteristics of the respondents in the two samples match very closely.

Table S1: Descriptive Statistics of Unweighted Survey, Target Sample Margins, and Weighted Survey

	Survey unweighted	2003–2013 VOX	Survey reweighted
Age	53.38	49.18	49.24
Female	0.50	0.53	0.53
Political Interest	3.31	2.87	2.88
Referendums	8.37	7.15	7.18
Education: 1	0.03	0.09	0.09
Education: 2	0.35	0.47	0.49
Education: 3	0.10	0.09	0.09
Education: 4	0.26	0.11	0.11
Education: 5	0.08	0.08	0.05
Education: 6	0.17	0.17	0.18
Employment	0.49	0.60	0.60

Table shows the descriptive statistics of the unweighted survey sample (Column 1), the VOX survey between 2003–2013 in the target municipalities that is our target sample (Column 2) and the reweighted survey sample (Column 3). Reweighting was performed using entropy balancing based on the following covariates: Age, Female (0/1), Political Interest (1–4), the number of referendums that respondents say they typically vote in assuming that there are 10 referendums per year (0-10), education (Education 1: compulsory schooling, Education 2: vocational training, Education 3: secondary schooling incl. *Matura*, Education 4: lower professional school, Education 5: higher professional school, Education 5: University degree) and Employment (0/1).

Student Sample

In addition to the main survey, we also conducted a similar experiment on a sample of Swiss undergraduate and graduate students as well as administrative and faculty staff of the University of Zurich. The participants were recruited between July 11, 2014 and August 3, 2014 via an email invite sent out to all students and University employees. One-third of all respondents were randomly assigned to answer the paired profiles conjont design with forced choice. $N = 652$

respondents completed this survey and form the basis for the student sample. A primary purpose of this additional experiment was to examine whether the results in the main experiment could also be replicated on a separate sample representing a very different population.

S4 Experimental Design

Attributes and Attribute Levels

Table S2 details the attributes and attribute levels used to generate the profiles. The attribute levels are randomized under the following two constraints to rule out illogical combinations: age $\geq$ years since arrival, and immigrants from Austria and Germany have a higher than "adequate" German language proficiency. The ordering of the attributes is fixed to match the typical leaflets as used in the actual naturalization referendums.

Table S2: Applicant Attributes and Attribute Levels
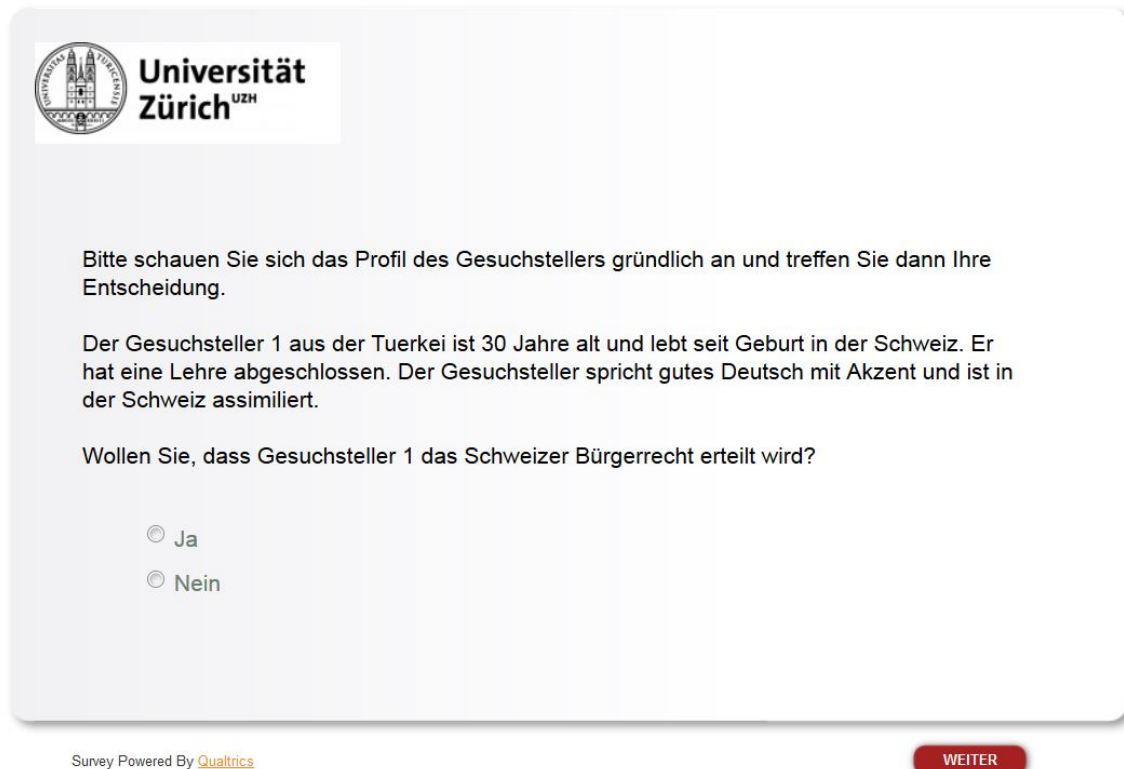
Attribute	Attribute Level
Gender $\in$	Male, Female
Origin $\in$	Germany, Austria, Netherlands, Italy, Turkey, Croatia, Former Yugoslavia, Bosnia and Herzegovina
Age $\in$	21 years, 30 years, 41 years, 55 years
Years since arrival $\in$	14 years, 20 years , 29 years, Born in CH
Education $\in$	Primary School, High School, University
German proficiency $\in$	"Adequate", "Good with accent", "Unaccented", "Swiss German"
Integration status $\in$	"Assimilated", "Integrated", "Indistinguishable", "Familiar with Swiss traditions"

Treatment Conditions: Five Survey Designs

For the core of the experiment, we asked participants to decide on naturalization applicants of immigrants. We randomly allocated respondents to five groups of equal size and presented each group with one of five survey formats, namely the *single vignette*, *paired vignette*, *single conjoint*, *paired conjoint*, and the *paired conjoint with forced choice*. In the following we describe each design. Each respondent completed ten choice tasks.

Figure S5 shows a screenshot from the *single vignette*. The design presents a single immigrant profile in the form of a short paragraph that describes the applicant with the attributes listed in the text and then respondents are asked to accept or reject the applicant.

Figure S5: Single Vignette

Figure shows *single vignette* in German. Attributes levels for Gender, Origin, Age, Years since arrival, Education, German proficiency and integration status are randomized subject to logical constraints. Attribute order is fixed. Respondents are asked to vote "yes" or "no" on each applicant.

Figure S6 shows a screenshot from the *paired vignette*. This design is similar to the single vignette except that two immigrant vignettes are presented below each other and then respondents are asked to accept or reject each of the two applicants.

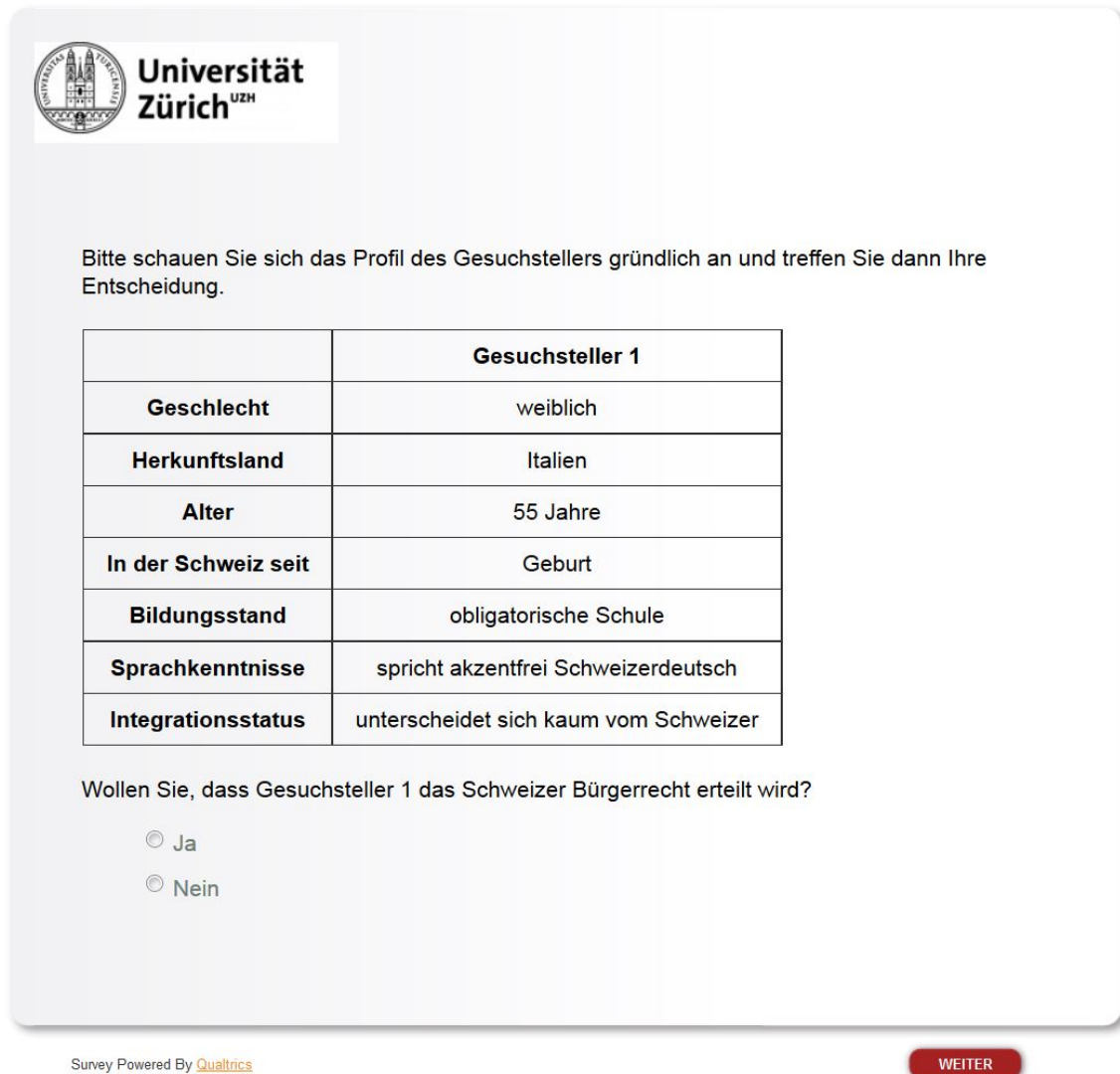Figure S6: Paired Vignette

Figure shows *paired vignette* in German. Attributes levels for Gender, Origin, Age, Years since arrival, Education, German proficiency and integration status are randomized subject to logical constraints. Attribute order is fixed. Respondents are asked to vote "yes" or "no" on each of the two applicants.

Figure S7 shows a screenshot from the *single conjoint*. This design presents a single immigrant profile in a conjoint table that resembles a CV with two columns. The first column lists the names of the attributes and the second column lists the attribute values. Again, respondents are asked to accept or reject the applicant.

Figure S7: Single conjoint
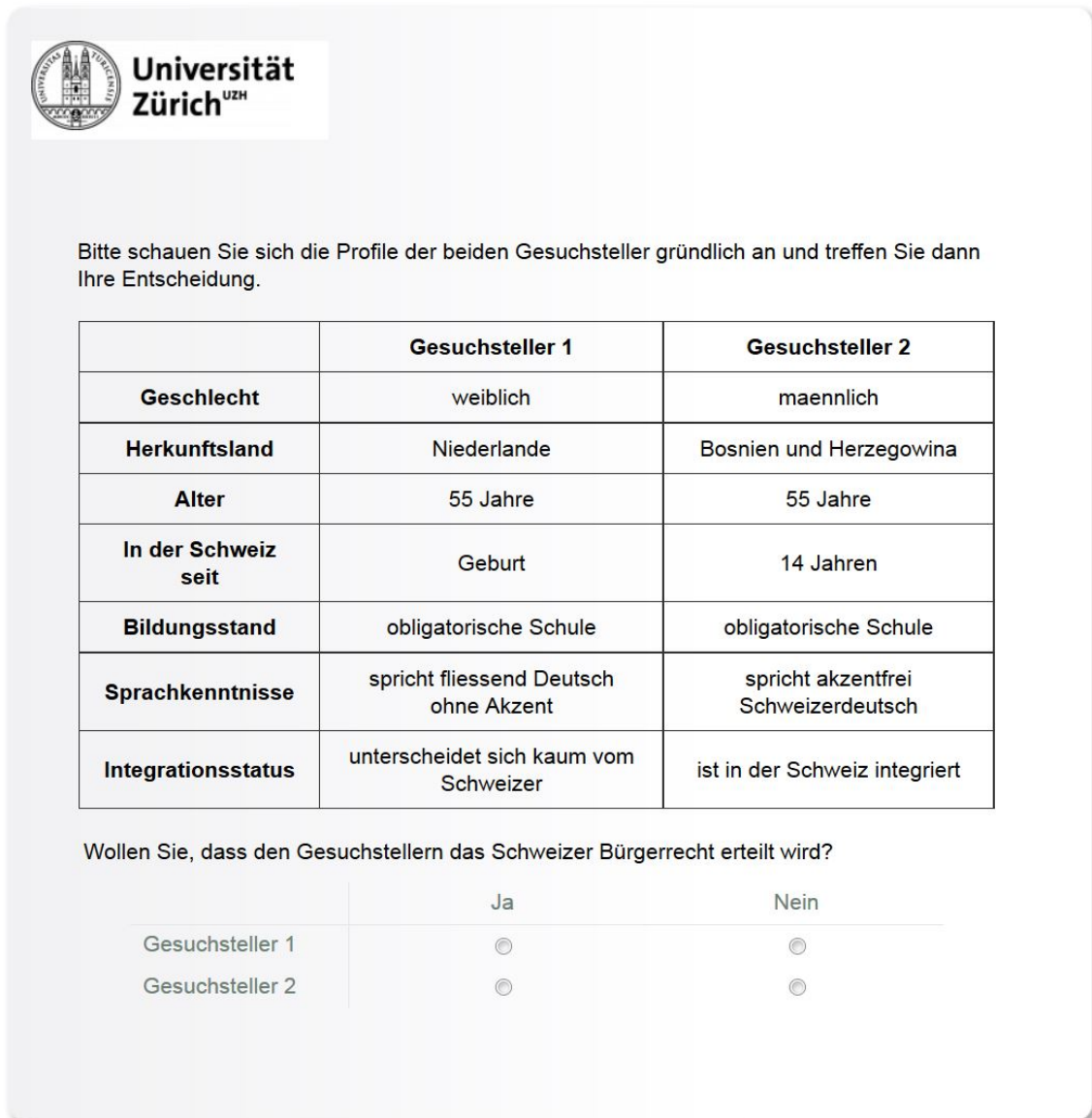
Figure shows *single conjoint* in German. Attributes levels for Gender, Origin, Age, Years since arrival, Education, German proficiency and integration status are randomized subject to logical constraints. Attribute order is fixed. Respondents are asked to vote "yes" or "no" on each applicant.

Figure S8 shows a screenshot from the *paired conjoint*. This design is similar to the single conjoint except that two immigrant profiles are presented next to each other in the conjoint table. Respondents are asked to accept or reject each of the two applicants.

Figure S8: Paired conjoint

Figure shows *paired conjoint* in German. Attributes levels for Gender, Origin, Age, Years since arrival, Education, German proficiency and integration status are randomized subject to logical constraints. Attribute order is fixed. Respondents are asked to vote "yes" or "no" on each of the two applicants.

Figure S9 shows a screenshot from the *paired conjoint with forced choice*. This design is identical to the paired conjoint except that respondents are asked to choose which of the two profiles they prefer for naturalization. In other words, respondents are forced to choose one of the two applicants and cannot accept or reject both.

Figure S9: Paired conjoint with forced choice

Figure shows *paired conjoint with forced choice* in German. Attributes levels for Gender, Origin, Age, Years since arrival, Education, German proficiency and integration status are randomized subject to logical constraints. Attribute order is fixed. Respondents are forced to choose one of the two applicants.

S5 Additional Results

This section reports additional analyses and robustness tests:

- Table S3 details the estimated effects of the applicant characteristics in actual and hypothetical naturalization referendums that are visualized in Figure 1 in the main text.

- Figure S10 shows the estimated differences in the effects of the applicant characteristics in the hypothetical and actual naturalization referendums. The estimates of the differences are generated based on a pooled dataset that combines the data from the hypothetical and actual naturalization referendums. In this pooled data we replicate the same model as in Table S3 and regress the rejection outcome on the attribute values, but also include a full set of indicator variables for the different survey experimental conditions plus the full set of interactions between these indicators and the attribute values. The coefficients on the interaction terms identify the differences in the estimates effects in hypothetical and actual naturalization referendums.

- Figure S11 and Table S4 replicate the main results based on the unweighted survey sample. The effects are very similar to the estimates based on the weighted sample displayed in Figure 1 and Table 1 in the main manuscript.

- Figure S12 and Table S5 replicate the main results but collapse the different country of origin indicators, following the coding of [1], into four roughly equal-sized categories: North West (Austria, Germany, Netherlands), South (Italy), Turkey, and Yugoslavia (Bosnia-Herzegovina, Croatia, and former Yugoslavia). Again, the results are very similar to the main results.

- Table S6 compares the estimated average rejection rate across the different survey designs to the behavioral benchmark. As discussed in the main text, most design underestimate the average rejection rate. The exception are the forced choice designs where the average rejection rate is by design fixed at 0.5.

Table S3: Attribute Effects in Actual and Hypothetical Naturalization Referendums

Condition	(1) Behavioral Benchmark	(2) Paired Conjoint	(3) Paired Conjoint Forced	(4) Paired Vignette	(5) Single Conjoint	(6) Single Vignette	(7) Paired Conjoint Forced Students
Gender:							
Male	0.0067	0.013	0.067*	0.00050	-0.0095	0.0088	0.027*
	(0.0067)	(0.014)	(0.029)	(0.015)	(0.019)	(0.013)	(0.012)
Origin:							
Germany	0.028	0.077	0.11*	0.077	0.00067	-0.0057	0.10**
	(0.023)	(0.041)	(0.055)	(0.047)	(0.059)	(0.031)	(0.031)
Austria	0.013	0.026	-0.020	-0.044	-0.021	-0.020	0.081**
	(0.038)	(0.035)	(0.041)	(0.036)	(0.051)	(0.044)	(0.031)
Italy	0.0030	0.0070	-0.011	-0.0085	-0.042	-0.037	0.0025
	(0.023)	(0.022)	(0.043)	(0.022)	(0.048)	(0.029)	(0.024)
Turkey	0.17**	0.16**	0.19**	0.087*	0.077*	0.036	0.044
	(0.028)	(0.039)	(0.046)	(0.035)	(0.031)	(0.030)	(0.024)
Bosnia & Herzegovina	0.19**	0.22**	0.19**	0.13**	0.097*	0.027	0.037
	(0.033)	(0.052)	(0.046)	(0.042)	(0.042)	(0.038)	(0.030)
Croatia	0.15**	0.12**	0.11*	0.10**	0.0016	0.046	0.024
	(0.027)	(0.040)	(0.057)	(0.035)	(0.043)	(0.039)	(0.029)
Yugoslavia	0.17**	0.13**	0.19**	0.094**	0.070	0.0015	0.015
	(0.026)	(0.039)	(0.053)	(0.033)	(0.037)	(0.039)	(0.029)
Age:							
30 Years Old	0.012*	0.00017	0.0027	-0.040	0.0035	0.045	-0.012
	(0.0057)	(0.028)	(0.039)	(0.028)	(0.024)	(0.030)	(0.018)
41 Years Old	0.015*	0.013	-0.011	0.018	0.11	0.059**	0.028
	(0.0067)	(0.023)	(0.040)	(0.037)	(0.073)	(0.019)	(0.019)
55 Years Old	0.0087	0.0026	0.059	0.024	0.039	0.046*	0.062**
	(0.0077)	(0.025)	(0.045)	(0.031)	(0.035)	(0.020)	(0.019)
Years Since Arrival:							
20 Years	-0.0018	-0.0034	-0.047	-0.028	-0.16**	-0.061*	-0.088**
	(0.0057)	(0.028)	(0.029)	(0.038)	(0.037)	(0.024)	(0.016)
29 Years	0.0090	-0.071**	-0.12**	-0.089*	-0.14*	-0.11**	-0.15**
	(0.012)	(0.024)	(0.029)	(0.040)	(0.065)	(0.031)	(0.018)
Born in Switzerland	-0.0074	-0.098**	-0.22**	-0.16**	-0.19**	-0.083**	-0.29**
	(0.012)	(0.026)	(0.037)	(0.034)	(0.058)	(0.028)	(0.017)
Education:							
Middle	-0.0091	-0.022	-0.095**	-0.048*	-0.039	-0.028	-0.12**
	(0.0074)	(0.028)	(0.023)	(0.024)	(0.023)	(0.021)	(0.015)
High	-0.032**	-0.071**	-0.056	-0.023	-0.030	-0.015	-0.17**
	(0.012)	(0.026)	(0.032)	(0.023)	(0.032)	(0.018)	(0.016)
Integration Status:							
Assimilated	-0.035	0.036	0.073*	0.042*	0.043	0.037	0.020
	(0.023)	(0.021)	(0.033)	(0.020)	(0.024)	(0.022)	(0.017)
Indistinguishable	-0.036*	-0.035	-0.035	0.038	0.00032	-0.043**	-0.092**
	(0.014)	(0.020)	(0.031)	(0.021)	(0.028)	(0.016)	(0.017)
Integrated	0.0016	0.027	0.0028	-0.00034	0.0090	0.0079	-0.048**
	(0.010)	(0.025)	(0.029)	(0.016)	(0.025)	(0.021)	(0.017)
German Proficiency:							
Good	0.0088	-0.042	-0.035	0.015	0.017	0.017	-0.087**
	(0.025)	(0.023)	(0.031)	(0.023)	(0.028)	(0.033)	(0.017)
Perfect	-0.015	-0.089**	-0.13**	-0.030	-0.029	-0.033	-0.20**
	(0.025)	(0.022)	(0.045)	(0.020)	(0.022)	(0.022)	(0.016)
Constant	0.37**	0.24**	0.57**	0.21**	0.23**	0.15**	0.82**
	(0.049)	(0.060)	(0.071)	(0.046)	(0.054)	(0.046)	(0.030)
Observations	1503	3910	3938	4274	2005	2173	6520

Ordinary least squares regression coefficients shown, with robust clustered standard errors in parentheses. Standard errors are clustered by the municipality (Model 1) or the respondents (Models 2-7) respectively. Model 1 is based on the actual naturalization referendums. Models 2-6 are based on our main survey and focus on the subsample of voters that is reweighted to match the margins of the Swiss post-referendum study VOX. Model 7 is based on the survey of the student sample. The reference categories for the various contrasts are: Gender: Female, Origin: Netherlands, Age: 21 Years, Years since Arrival: 14 Years, Education: Low, Integration Status: Traditions, German Proficiency: Adequate. Model 1 for the actual naturalization referenda also includes municipality and period fixed effects.

Figure S10: Differences in Effects of Applicant Attributes: Survey versus Behavioral Estimates

Figure shows point estimates (dots) and corresponding, cluster-robust 95 % confidence intervals (horizontal lines) from ordinary least squares regressions that identify the differences in the estimated effects in the survey conditions and the behavioral benchmark. The dots on the zero line without confidence intervals denote the reference category for each applicant attribute.

Figure S11: Effects of Applicant Attributes on Opposition to Naturalization Request (Unweighted Survey Sample)
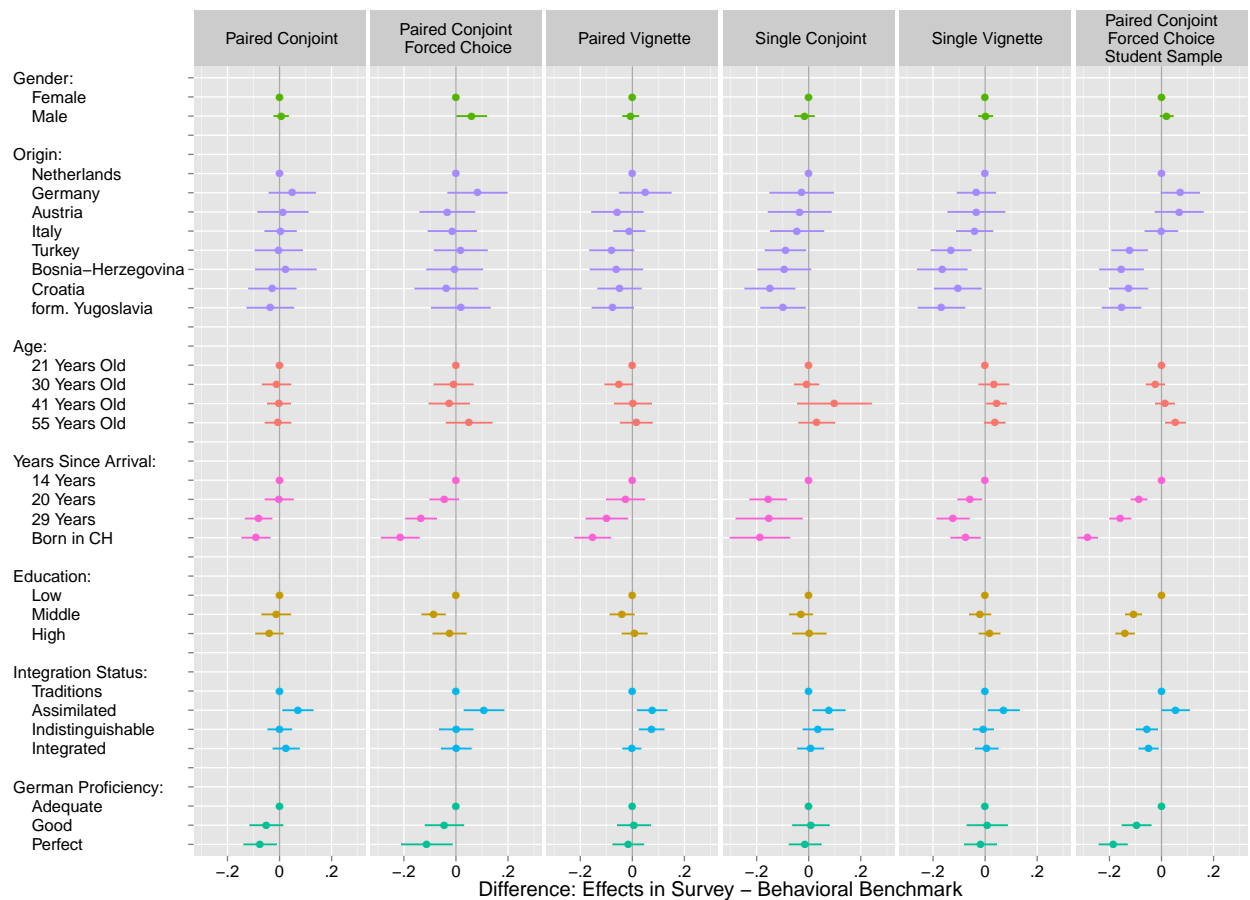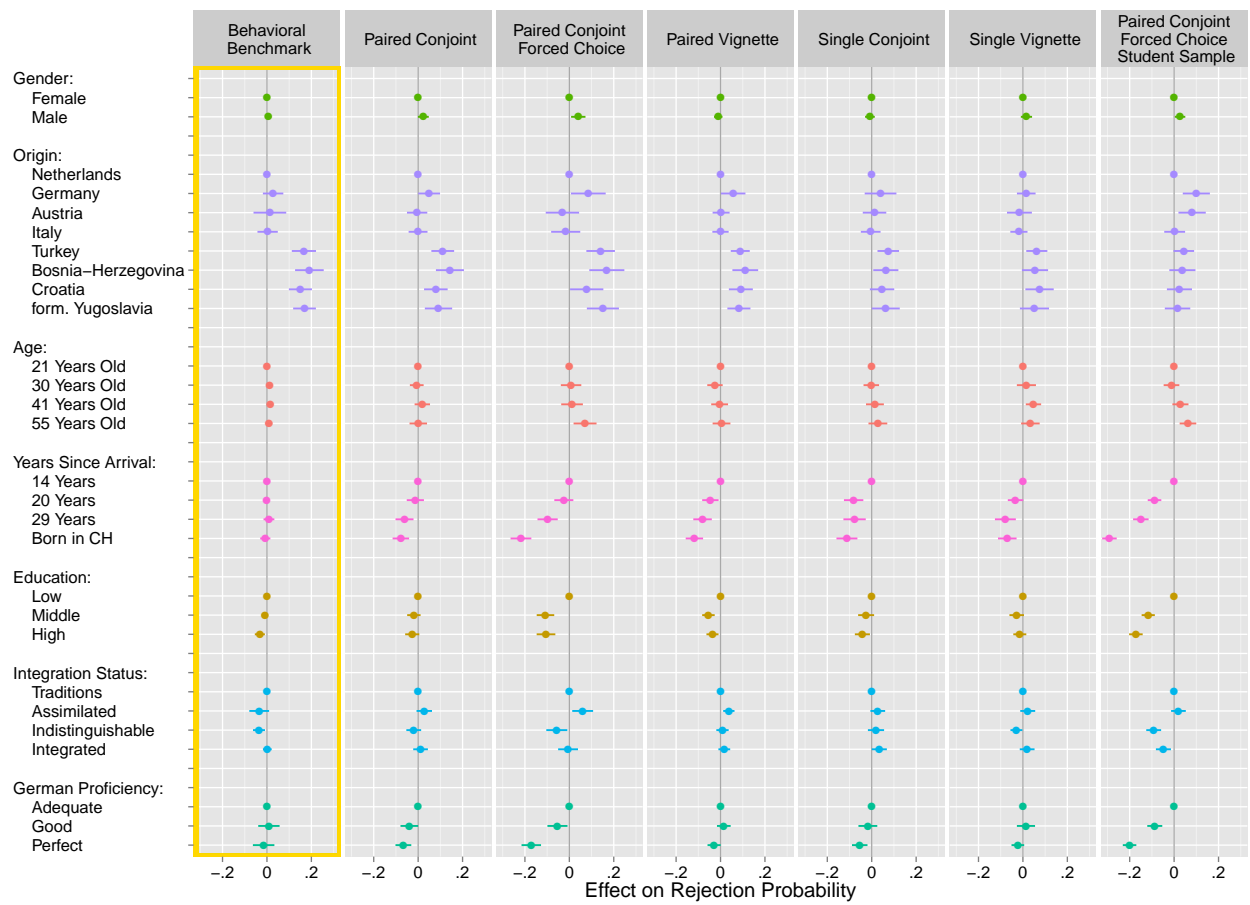

Figure shows point estimates (dots) and corresponding, cluster-robust 95 % confidence intervals (horizontal lines) from ordinary least squares regressions. The dots on the zero line without confidence intervals denote the reference category for each applicant attribute.

Table S4: Differences in Effects of Applicant Attributes: Survey versus Behavioral Estimates (Unweighted Survey Sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Absolute Differences			Sig. Diffs		Joint		
Design:	mean	median	max	raw	adj	F-test	$Cor(Y, \hat{Y})$	$Cor(\hat{Y}_b, \hat{Y}_s)$
Paired Conjoint	0.02	0.01	0.08	3/21	1/21	2.04	0.41	0.70
Paired Conjoint, FC	0.04	0.02	0.21	7/21	4/21	10.62	0.32	0.55
Paired Vignette	0.03	0.02	0.11	9/21	2/21	4.35	0.35	0.60
Single Conjoint	0.04	0.02	0.13	9/21	3/21	2.94	0.33	0.57
Single Vignette	0.03	0.02	0.14	6/21	2/21	2.82	0.35	0.60
Paired Conjoint, FC (Students)	0.07	0.06	0.28	14/21	11/21	26.69	0.13	0.23
Behavioral							0.58	

Table reports measures of performance for each survey design based on the unweighted sample of voters. Column 1–3 display the mean, median, and maximum of the absolute differences from the behavioral benchmark across the 21 attribute effects. Column 4 shows the total number of differences from the benchmark estimates that are statistically different from zero at the .05 significance level. Column 5 presents the same metric but with the Bonferroni correction. Column 6 presents an F-statistic for the hypothesis test against the joint null of no difference between the effects in the behavioral benchmark and each survey design. Column 7 presents the bivariate correlation between observed shares of rejection votes and the predicted rejection probabilities. Column 8 presents the bivariate correlation between the predicted rejection probabilities based on the survey estimates and the fitted rejection rates in the behavioral regression. See main text for further details on the procedure used to generate columns 7 and 8.

Figure S12: Effects of Applicant Attributes on Opposition to Naturalization Request (Aggregated Origin Groups)
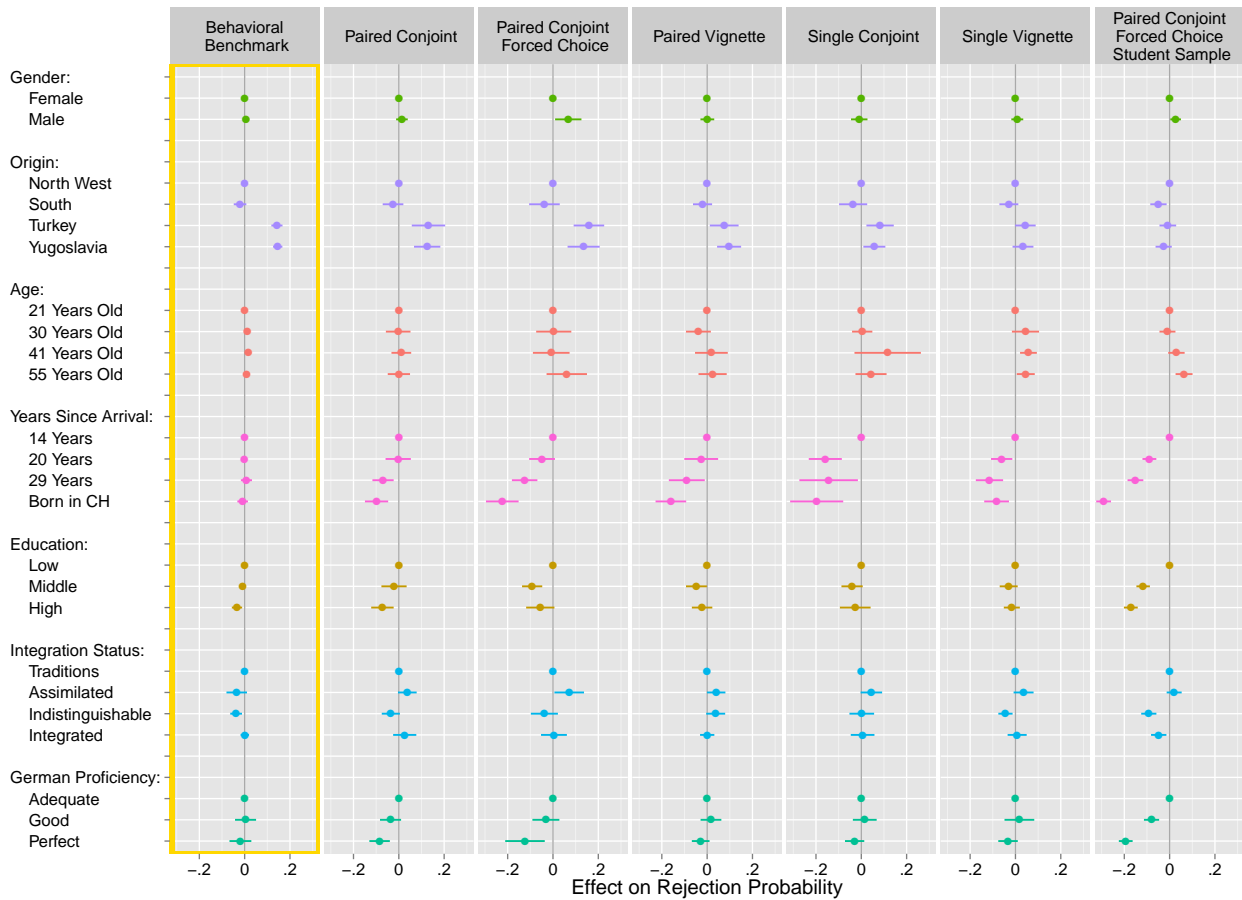

Figure shows point estimates (dots) and corresponding, cluster-robust 95 % confidence intervals (horizontal lines) from ordinary least squares regressions. The dots on the zero line without confidence intervals denote the reference category for each applicant attribute.

Table S5: Differences in Effects of Applicant Attributes: Survey versus Behavioral Estimates (Aggregated Origin Groups)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Absolute Differences			Sig. Diffs		Joint		
Design:	mean	median	max	raw	adj	F-test	$\text{Cor}(Y, \hat{Y})$	$\text{Cor}(\hat{Y}_b, \hat{Y}_s)$
Paired Conjoint	0.02	0.01	0.09	4/17	1/17	2.29	0.47	0.80
Paired Conjoint, FC	0.04	0.02	0.21	6/17	3/17	11.88	0.37	0.64
Paired Vignette	0.03	0.01	0.15	4/17	2/17	3.76	0.34	0.59
Single Conjoint	0.04	0.01	0.19	5/17	3/17	4.96	0.28	0.47
Single Vignette	0.03	0.01	0.12	7/17	3/17	3.94	0.28	0.49
Paired Conjoint, FC (Students)	0.07	0.05	0.28	12/17	8/17	31.15	0.17	0.30
Behavioral							0.58	

Table reports measures of performance for each survey design based on the weighted sample of voters and based on the aggregated origin groups. Column 1–3 display the mean, median, and maximum of the absolute differences from the behavioral benchmark across the 21 attribute effects. Column 4 shows the total number of differences from the benchmark estimates that are statistically different from zero at the .05 significance level. Column 5 presents the same metric but with the Bonferroni correction. Column 6 presents an F-statistic for the hypothesis test against the joint null of no difference between the effects in the behavioral benchmark and each survey design. Column 7 presents the bivariate correlation between observed shares of rejection votes and the predicted rejection probabilities based on the survey estimates. Column 8 presents the bivariate correlation between the predicted rejection probabilities based on the survey estimates and the fitted rejection rates in the behavioral regression. See main text for further details on the procedure used to generate columns 7 and 8.

Table S6: Estimated Average Rejection Rate for the Applicants with Naturalization Referendums

	Estimated Average Rejection Rate
Behavioral Benchmark	.37
Paired Conjoint	.21
Paired Conjoint Forced	.49
Paired Vignette	.17
Single Conjoint	.12
Single Vignette	.10
Paired Conjoint Forced Students	.47

Table shows the estimated average rejection rate for the applicants with naturalization referendums. For the behavioral benchmark the rejection rate is simply the average proportion voting "no" in the referendum sample. For each survey condition we predict the rejection probability for the applicants in the referendum sample by taking their characteristics and multiplying them with the coefficients estimated from the survey respondents and then take the average of these predicted values. For observations with missing attribute information in the behavioral data, we impute missing values with their observed mean levels.

S6   Survey Engagement

This section analyses the differences in respondents' survey engagement across the different designs and thereby offers at least suggestive evidence for one particular causal pathway that runs through survey engagement, and explains why the paired designs produce better estimate of attribute effects than the single profile design.

Figure S13 shows that respondents in the paired and single profile conditions perceived no significant difference in the length of the survey, even though the actual response time was about 60% longer. Median response time used to complete the 10 decision tasks was 245 seconds for the *paired conjoint*, 291 seconds for the *paired conjoint with forced choice*, 253 seconds for the *paired vignette*, 166 seconds for the *single vignette*, and 153 seconds for the *single conjoint*.

Figure S14 shows that respondents perceived no significant difference in the difficulty of the survey, even though respondents in the paired profile conditions evaluated twice as many applicant profiles.

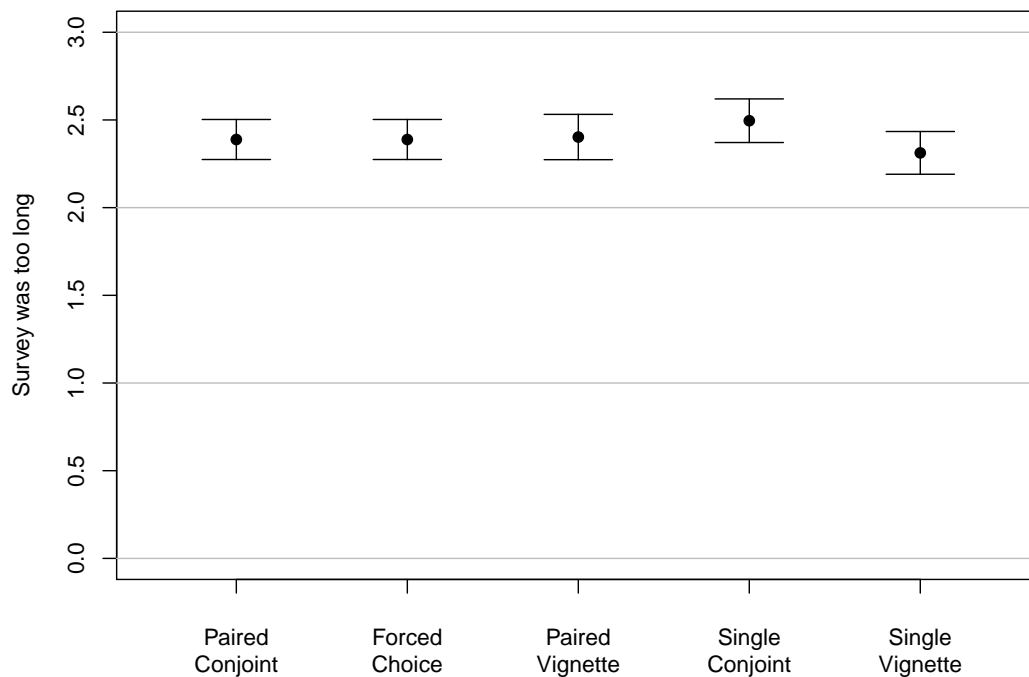Figure S13: Perceived Survey Length Across Survey Designs

Figure shows estimated means and corresponding 95 % confidence intervals for perceived of survey length. At the end of the survey, respondents were asked if they agree that the survey was too long (4: completely agree, 3: agree, 2: neither, 1: disagree, 0: completely disagree)

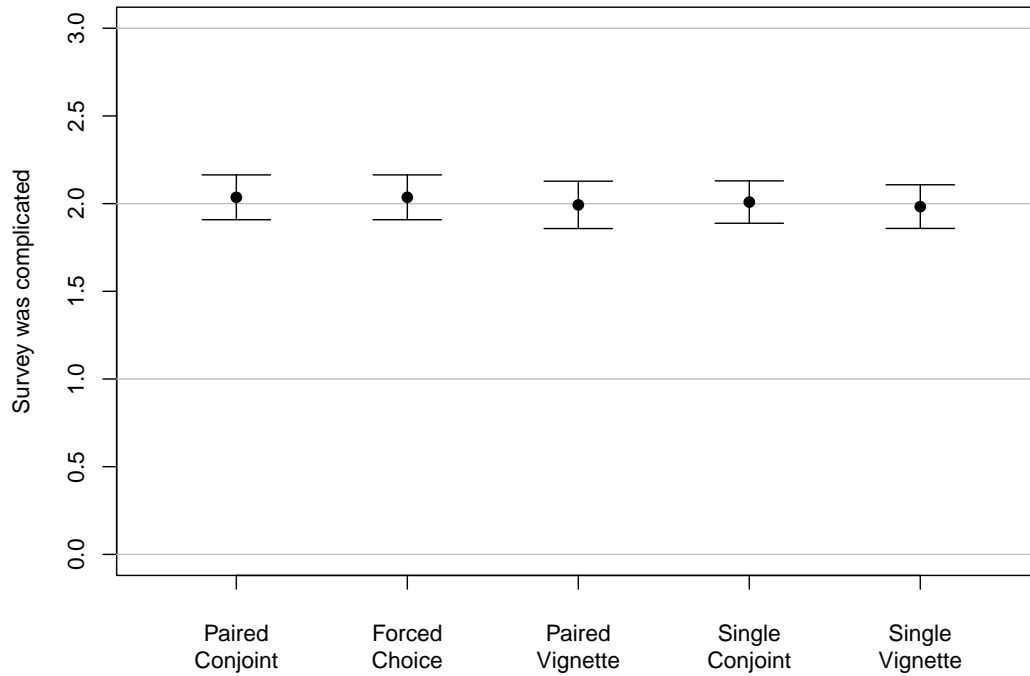Figure S14: Perceived Survey Difficulty Across Survey Designs

Figure shows estimated means and corresponding 95 % confidence intervals for perceived survey difficulty. At the end of the survey, respondents were asked if they agree that the survey was "complicated" (4: completely agree, 3: agree, 2: neither, 1: disagree, 0: completely disagree)

References

[1] Hainmueller J, Hangartner D (2013) Who gets a swiss passport? a natural experiment in immigrant discrimination. *American Political Science Review* 107:159–187.

[2] Hainmueller J, Hopkins D (Forthcoming) The hidden american immigration consensus: A conjoint analysis of attitudes toward immigrants. *American Journal of Political Science* p DOI: 10.1111/ajps.12138.

[3] Hainmueller J (2011) Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* 20:25–46.