In Stars We Trust – A Note on Reputation Portability Between Digital Platforms

Maik Hesse, Timm Teubner, Marc T. P. Adam

Business & Information Systems Engineering (2021)

Appendix (available online via http://link.springer.com)

Appendix A – Stimulus Material

Imagine that you would like to book a stay via the (hypothetical) Internet platform **happystay.com**. On this platform, private hosts rent out their apartments and rooms (very similar to Airbnb). You have already identified a potential host and, in the following, consider the host's profile page. Note that you will see a very stylized form of this page (i.e., with a blurred profile photo).

Based on the host's (star) ratings, we will ask you to state how likely you are to book and how trustworthy you consider the host to be.

- 1: Note, however, that the host has not collected any ratings yet.
- 2: Note that the host has already collected ratings from previous guests on happystay.com.
- 3: Note that the host has not collected any ratings on happystay.com yet but has imported previous ratings from her/his Airbnb profile.
- 4: Note that in addition to her/his ratings on happystay.com, the host has imported previous ratings from her/his Airbnb profile.

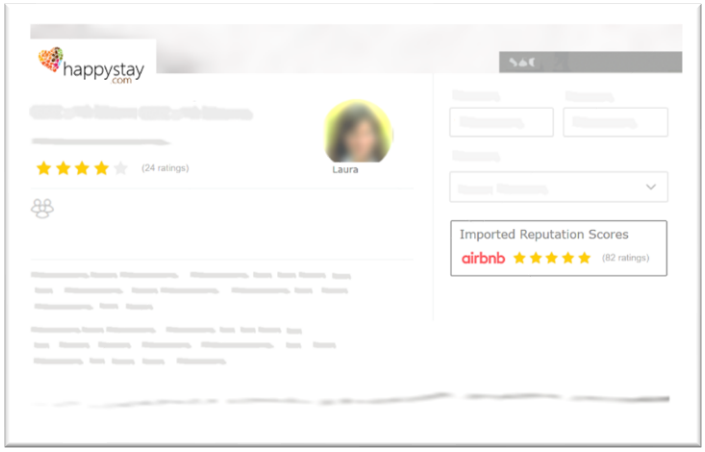

Figure A1 Landing page for online experiment participants

Names: (equally distributed; randomly drawn and allocated to blurred female or male image)

Alex, Emma, Laura, Lisa, Lucas, Michael, Paul, Philipp, Pia, Sarah, Sophia, Tim

Images:

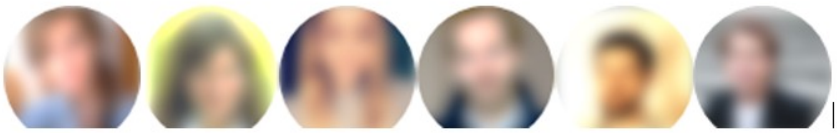

Figure A2 Blurred profile images of complementors on platform

Appendix B – Treatment Design

Table B1 shows the experiment's treatment matrix. As can be seen, when no rating is available in the 2x2 between subjects design, the condition has been coded as "NA". The participants' allocation across cases (1) to (4) is indicated by the color-coding, namely case (1) (no rating; 24 observations), case (2) (on-site only; 81 observations), case (3) (imported only, 83 observations), and case (4) (both ratings; imported and/or on-site rating ranges from 1.0 to 5.0; 305 observations).

Table B1 Treatment matrix (i.e., participants' allocation across cases 1 to 4)

		Imported rating										
		1.0	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**	**5.0**	**NA**	**∑**
On-site rating	**1.0**	2	2	2	2	2	3	3	2	3	7	**28**
	1.5	2	2	2	2	2	3	3	3	3	7	**29**
	2.0	2	2	2	2	2	4	4	4	4	6	**32**
	2.5	2	2	2	2	2	4	4	3	4	7	**32**
	3.0	2	2	2	2	2	4	4	4	4	7	**33**
	3.5	4	4	3	4	3	7	7	7	7	11	**57**
	4.0	4	4	4	4	4	6	6	7	7	13	**59**
	4.5	4	4	3	4	4	6	6	7	7	11	**56**
	5.0	4	4	4	4	4	7	7	7	7	12	**60**
	NA	6	6	7	6	7	13	12	13	13	24	**107**
	∑	**32**	**32**	**31**	**32**	**32**	**57**	**56**	**57**	**59**	**105**	**493**

Table B2 reports descriptive statistics (i.e., mean / std. dev.) of participant-specific control variables. All controls were evenly distributed, and no systematic differences existed for participants' key characteristics across treatments (i.e., results were not driven by confounding factors).

Table B2 Balance of participant-specific controls across cases

		Case				
Control	**Scale**	1 (no rating)	2 (on-site only)	3 (imported only)	4 (both ratings)	**All cases**
Disposition to Trust	1-11	6.1 (2.5)	6.3 (2.6)	5.8 (2.5)	5.9 (2.6)	6.0 (2.6)
Trust in Airbnb	1-11	7.4 (2.2)	6.9 (2.5)	7.1 (2.4)	7.0 (2.5)	7.1 (2.5)
Risk-Affinity	1-11	3.8 (2.3)	4.4 (2.7)	4.4 (2.6)	4.5 (2.7)	4.4 (2.7)
Online Shopping Experience	1-5	3.0 (1.2)	2.5 (1.3)	2.9 (1.3)	2.9 (1.3)	2.9 (1.3)
Age	18-70	33.1 (9.5)	32.8 (12.0)	34.7 (11.9)	35.9 (12.2)	35.0 (12.0)
Gender Female	{0, 1}	66%	59%	57%	53%	56%

Note: Mean values reported (standard errors in parentheses)

Table B3 shows that no significant differences occurred for the number of reviews that were displayed in the different groups. Please also refer to **Appendix D – Supplemental Analysis**, in which we additionally investigate the interaction of rating score and the number of reviews.

Table B3 Number of reviews across treatment variations

		Number of ratings				
Type of rating	**n**	Minimum	Maximum	Mean	St. Dev.	Median
On-site	386	1	89	22.4	28.1	9
Imported	388	1	89	24.4	29.9	9

Appendix C – Questionnaire and Measurement Instruments

Table C1 Instrument variables (measured on 11-point Likert scales)

Control	Question	Source
Trust in Complementor	I would consider this host as trustworthy.	(adapted from Gefen, 2002)
Risk-Affinity	In general, I am very willing to take risks.	(adapted from Dohmen et al., 2011)
Disposition to Trust	I generally trust other people.	(adapted from Gefen, 2000)
Trust in Airbnb	I think that Airbnb is a trustworthy company.	(adapted from Gefen, 2002)

Appendix D – User Perceptions of Rating Distributions

Note the derived thresholds in this study should not been seen as absolute or definitive but rather as the concrete representation of the more general idea that there will emerge *some* strategic decision boundaries for reputation import. Of course, these values will depend on the specific sample of study participants, the studied platform(s), as well as participants' implicit perceptions of how rating scores are distributed. To allow for some degree of generalizability, we conducted an additional survey with a sample of n=441 participants, and asked them to provide their estimates on the rating score distribution across ten platforms (including the hypothetical Happystay.com platform used in the paper). Participants entered percentage values for each rating on the scale from 1.0 to 5.0 stars (in steps of 0.5 stars). These values were required to add up to 100 percent. We used a between-subjects design, hence every participant provided only one distribution. On average, this yielded ~44 participants per platform. The results of this inquiry are shown in Figure D1.

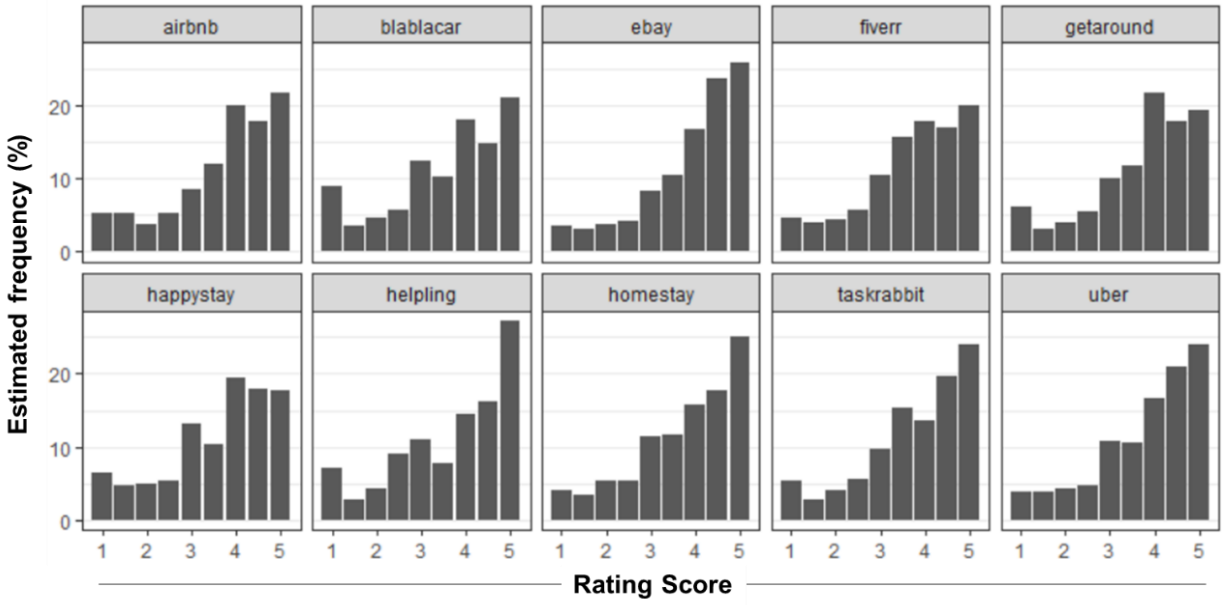

Figure D1 Users' perceived rating distributions across a set of platforms

The data show that these "perceived rating distributions" are quite similar across platforms. Specifically, we find an average correlation between the distributions of $\bar{r}$ =.928. All of the 10×9/2 = 45 pairwise correlations exhibit significance at the 1% level (in fact, the "weakest" correlation coefficient is r_{min} = .820; *p*=.0067). Importantly, across all platforms, users seem to be aware of some skewness towards good ratings, but still markedly underestimate the actual degree of skewness (Hesse et al., 2020; Schoenmueller et al., 2018). For our study's findings, this perceived similarity supports our reasoning in the sense that a rating of a certain value from platform A does not mean something completely different compared to a similar rating on platform B, providing robustness to our findings.

Note that all participants for this additional survey were required to have at least some experience with "online shopping." Moreover, as we employed the online platform Prolific.co to recruit the sample, it can be assumed that participants had reasonable levels of online literacy and savviness.

Appendix E – Trust-maximizing Strategies

As a complementary analysis, consider a complementor who has *both* an on-site and an incumbent rating and is making a decision about which of these rating(s) to display. While, certainly, complementors would (and should) not have a choice to "drag & drop" their reputation at will, we consider these hypothetical scenarios to investigate *all* possible combinations and the resulting theoretical implications. Analyzing the resulting decision thresholds from all pairwise comparisons yields the full strategy for obtaining the highest level of trust perceptions. Figure E1 shows all six resulting thresholds for the pair-wise case comparisons and highlights the respective *trust-maximizing strategies.* Depending on the specific ratings, the complementor may not display any rating, display either the imported *or* the on-site rating only, or both ratings simultaneously. Irrelevant (sections of) boundaries are shown as dotted lines.

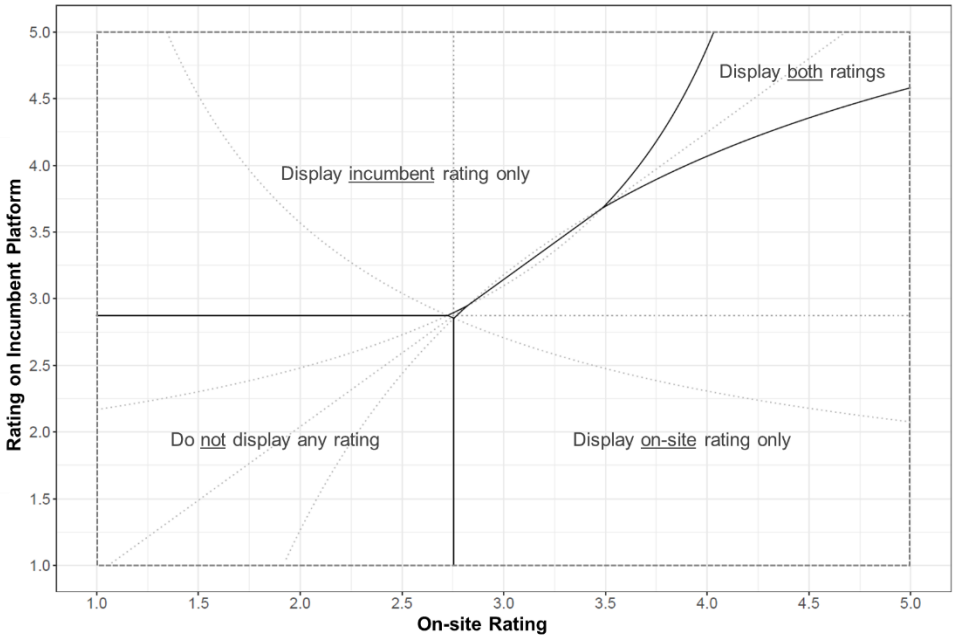

Figure E1 Trust-maximizing rating display (given that both ratings are available)

Appendix F – Supplemental Analysis

Full control variable analysis—Table F1 reports the full *control variable analysis* of all used variables across the four cases. For each case 1 to 3 we show the original model results next to an extended model comprising all relevant controls per case (i.e., participant-specific controls and number of reviews; Models 1.1/2, 2.1/2, 3.1/2). Additionally, we repeated the analyses for case 4 including the original model (i.e., Model 4.1), including all controls (Model 4.2), and a model without interaction (Model 4.3; to check for significance when controlling for both ratings). Please note that for this analysis we introduced mean-centered control variables to account for the fact that the original coefficients are determined by these controls and can only be interpreted in a meaningful manner when applied to "the average population" (which is achieved by mean-centering). In particular, it can be seen that in case 1, the coefficient α_1, calculated as .350 in our model, is only slightly affected when including all (mean-centered) controls (i.e., equating to .376), despite the statistical significance of some of the control variables. The same holds for all other relevant coefficients from the "trust function", which can be checked in the control models 2.2, 3.2, and 4.2 in Table F1 – they do not differ substantially in their sign, significance, or magnitude.

Interaction of rating score and number of reviews—Table F2 reports the results of a robustness check for the interaction of *rating score* and *number of reviews*. For that purpose, we calculated a series of multivariate OLS regression models to analyze the interaction effect across the three cases where at least one rating was visible (i.e., cases 2, 3, and 4). Neither in case 2 nor 3 (onsite or imported rating only; Models 2a/b/c, 3a/b/c) the number of reviews has any significant impact. Both when controlling only for the number of reviews and when additionally controlling for the interaction with the respective rating. Similarly, in case 4 (both ratings; Models 4a/b/c) the number of reviews does have a significant impact (Model 4b; b=.001, p<.05), albeit negligibly small – it would take 1,000 additional reviews to increase the evaluation of trust by one unit on the 1-7 Likert scale. Additionally, this significant effect vanishes when also controlling for the interaction(s) between rating and number of reviews (Model 4c). In conclusion, the controls barely have any influence on coefficients – neither in terms of magnitude nor significance.

Table F1 OLS regressions; extension by (mean-centered) control variables; standard errors in parentheses

	DV: **Trust in Complementor**								
	Model 1.1 (Case 1)	**Model 1.2** (Case 1)	**Model 2.1** (Case 2)	**Model 2.2** (Case 2)	**Model 3.1** (Case 3)	**Model 3.2** (Case 3)	**Model 4.1** (Case 4)	**Model 4.2** (Case 4)	**Model 4.3** (Case 4)
Constant	0.350 **	0.376 ***	-0.004	0.081	0.016	0.021	0.259 *	0.290 **	-0.010
	(0.038)	(0.037)	(0.075)	(0.062)	(0.082)	(0.071)	(0.116)	(0.099)	(0.057)
On-site Rating			0.128 ***	0.106 ***			-0.029	-0.017	0.050 ***
			(0.021)	(0.017)			(0.032)	(0.027)	(0.012)
Imported Rating					0.116 ***	0.116 ***	-0.006	-0.005	0.076 ***
					(0.023)	(0.020)	(0.033)	(0.028)	(0.012)
On-site × Imported							0.024 **	0.017 *	
							(0.009)	(0.008)	
Disposition to Trust		0.018		0.005		0.039 **		0.029 ***	
		(0.014)		(0.009)		(0.012)		(0.005)	
Risk-Affinity		0.043 *		0.028 **		0.023 *		0.031 ***	
		(0.017)		(0.009)		(0.010)		(0.005)	
Trust in Airbnb		0.001		0.021 *		0.013		0.014 *	
		(0.017)		(0.010)		(0.012)		(0.006)	
Online Shopping Experience		-0.004		0.058 **		-0.032		-0.007	
		(0.029)		(0.017)		(0.020)		(0.009)	
Age		-0.0004		-0.004 *		0.001		0.002	
		(0.004)		(0.002)		(0.002)		(0.001)	
Participant Female		-0.013		-0.049		0.062		-0.023	
		(0.081)		(0.044)		(0.052)		(0.025)	
#Reviews (ons.)				-0.0003				0.001	
				(0.001)				(0.0004)	
#Reviews (imp.)						0.001		-0.001	
						(0.001)		(0.0004)	
Observations	24	24	81	81	83	81[1)]	305	300[1)]	305
R^2	-	0.476	0.317	0.625	0.246	0.534	0.189	0.472	0.170
Adjusted R^2	-	0.292	0.308	0.583	0.237	0.482	0.181	0.451	0.165
Res. Std. Error	0.184 (df = 23)	0.155 (df = 17)	0.244 (df = 79)	0.189 (df = 72)	0.257 (df = 81)	0.214 (df = 72)	0.255 (df = 301)	0.209 (df = 288)	0.258 (df = 302)
F Statistic		2.579 (df = 6; 17)	36.646*** (df = 1; 79)	14.988*** (df = 8; 72)	26.471*** (df = 1; 81)	10.309*** (df = 8; 72)	23.389*** (df = 3; 301)	23.364*** (df = 11; 288)	30.978*** (df = 2; 302)

1) Age information missing for some participants; observations excluded from analysis

*** p<.001; ** p<.01; * p<.05

Table F2 OLS regressions; interaction of rating score and number of reviews; standard errors in parentheses

	DV: **Trust in Complementor**								
	Model 2a (Case 2)	**Model 2b** (Case 2)	**Model 2c** (Case 2)	**Model 3a** (Case 3)	**Model 3b** (Case 3)	**Model 3c** (Case 3)	**Model 4a** (Case 4)	**Model 4b** (Case 4)	**Model 4c** (Case 4)
Constant	-0.004	0.002	0.050	0.016	0.010	-0.044	0.259 *	0.217	0.214
	(0.075)	(0.078)	(0.095)	(0.082)	(0.085)	(0.109)	(0.116)	(0.119)	(0.129)
On-site Rating	0.128 **	0.128 ***	0.114 ***				-0.029	-0.022	-0.024
	(.012)	(0.021)	(0.027)				(0.032)	(0.032)	(0.035)
Imported Rating				0.116 ***	0.116 ***	0.131 ***	-0.006	0.001	0.003
				(0.023)	(0.023)	(0.030)	(0.033)	(0.033)	(0.035)
On-site × Imported							0.024 **	0.022 *	0.022 *
							(0.009)	(0.009)	(0.009)
#Reviews (on-site)		-0.0003	-0.002					0.001 *	0.001
		(0.001)	(0.003)					(0.001)	(0.001)
#Reviews (imported)					0.0003	0.002		-0.0004	0.00002
					(0.001)	(0.003)		(0.0005)	(0.001)
Rating × #Reviews (on-site)			0.001						0.0001
			(0.001)						(0.0004)
Rating × #Reviews (imported)						-0.001			-0.0001
						(0.001)			(0.0004)
Observations	81	81	81	83	83	83	305	305	305
R^2	0.317	0.318	0.324	0.246	0.247	0.253	0.189	0.204	0.204
Adjusted R^2	0.308	0.300	0.298	0.237	0.228	0.224	0.181	0.191	0.186
Res. Std. Error	0.244 (df = 79)	0.245 (df = 78)	0.246 (df = 77)	0.257 (df = 81)	0.258 (df = 80)	0.259 (df = 79)	0.255 (df = 301)	0.254 (df = 299)	0.255 (df = 297)
F Statistic	36.646*** (df = 1; 79)	18.152*** (df = 2; 78)	12.323*** (df = 3; 77)	26.471*** (df = 1; 81)	13.136*** (df = 2; 80)	8.910*** (df = 3; 79)	23.389*** (df = 3; 301)	15.329*** (df = 5; 299)	10.896*** (df = 7; 297)

*** p<.001; ** p<.01; * p<.05

Gender effects—We first analyzed mean trust values (on a scale of 0 to 1) to check for a general tendency towards differences in the assigned value of trust scores among all possible host/guest combinations. Table F3 shows these mean values for all tested combinations of male/female hosts (i.e., profile pages) as well as male/female guests (i.e., survey participants). As we can see, evaluations of trust are generally higher for female hosts than they are for males and, further, the combination female guest/participant and male host exhibits the lowest overall trust average.

Table F3 Number of observations and average trust scores per host/guest combination

Case	Observations	Trust Score
Male host, male guest	122	.393
Male host, female guest	131	.379
Female host, male guest	97	.431
Female host, female guest	143	.436
All combinations	**493**	**.410**

To better understand these results, we introduced binary dummy variables for *Female Host* and *Female Guest* (i.e., participant) to check if above "effects" can be validated statistically. As we see in Table F4, a female host has a positive (albeit small) impact on participants' evaluations of trust (Model 1) – irrespective of the prospective guest's gender. However, the effect itself – as any of the investigated gender control variables – is not statistically significant. Similarly, female guests do not exhibit a significantly different evaluation of trust in hosts (Models 2/3; also, when controlling for host gender). Further, we checked the interaction term for the combination female hosts / female guests (Model 4). As expected, this validates our results on the trust scores as reported in Table F3, however, the gender effects have no significant impact on our model (note: this also holds for the combination female guest/ male host, which induces negative effects on trust, but these are not statistically significant as per Models 3/4). As a final check, we also controlled for the influence of the same gender on trust (i.e., a female guest evaluating a female host or vice versa with males; Model 5). Again, the effect is negligible in size and not significant (b=.015, p=.55). Table F4 summarizes the results and we conclude that there is no systematic gender bias in our data.

Table F4 OLS regressions; effects of gender (host/participant) on assigned trust scores

	DV: **Trust in Complementor**				
	Model (1) (all cases)	**Model (2)** (all cases)	**Model (3)** (all cases)	**Model (4)** (all cases)	**Model (5)** (all cases)
Constant	0.386 ***	0.410 ***	0.389 ***	0.393 ***	0.401 ***
	(0.018)	(0.019)	(0.022)	(0.025)	(0.019)
Female Host	0.048		0.048	0.037	
	(0.025)		(0.025)	(0.038)	
Female Guest		-0.001	-0.005	-0.014	
		(0.026)	(0.026)	(0.035)	
Interaction				0.019	
				(0.051)	
Same Gender					0.015
					(0.025)
Observations	493	493	493	493	493
Adjusted R^2	0.005	-0.002	0.003	0.002	-0.001

*** p<.001; ** p<.01; * p<.05

Last, we also checked for the effect of host names and profile pictures on assigned trust scores. Figure F1 reports the mean assigned trust values (incl. their standard deviation) per either host name or blurred profile image as previously introduced in **Appendix A – Stimulus Material**. The analysis indicates the tendency that, on average, some female host names (e.g., Emma, Laura) exhibit higher trust scores than some of their male counterparts (e.g., Philipp, Paul, Michael). Similarly, blurred profile images with a female contour exhibit higher trust scores than male blurred profile pictures. However, neither of the observed effects is statistically significant at the 5%-level (p=.068 for names; p=.406 for images).

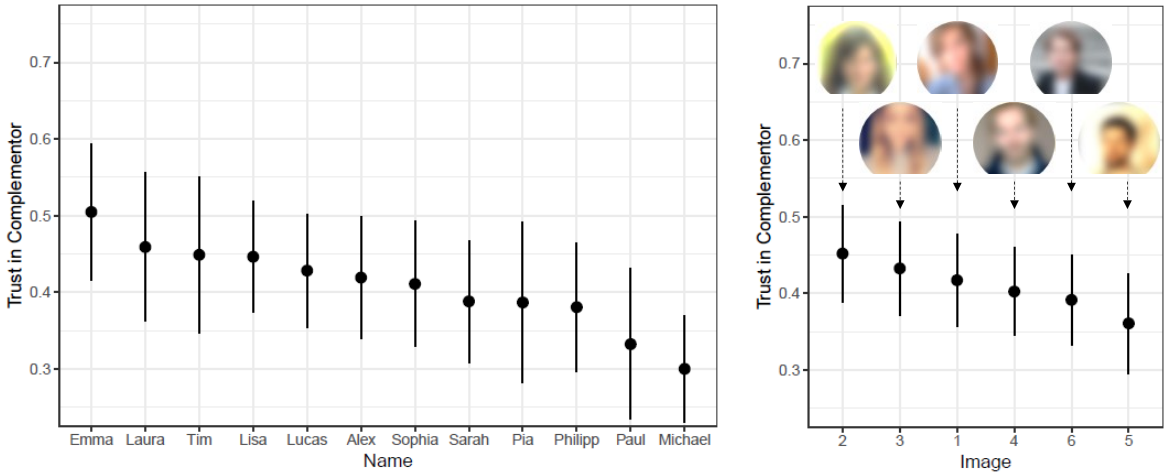

Figure F1 Mean trust scores per host name and profile image (error bars show 95% confidence intervals)

References

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: Measurement, determinants, and behavioral consequences. J. Eur. Econ. Assoc. 9, 522–550. https://doi.org/10.1111/j.1542-4774.2011.01015.x

Gefen, D., 2002. Reflections on the Dimensions of Trust and Trustworthiness among Online Consumers. Data Base Adv. Inf. Syst. 33, 38–53. https://doi.org/10.1145/569905.569910

Gefen, D., 2000. E-commerce: The role of familiarity and trust. Omega 28, 725–737. https://doi.org/10.1016/S0305-0483(00)00021-9

Hesse, M., Dann, D., Braesemann, F., Teubner, T., 2020. Understanding the Platform Economy: Signals, Trust, and Social Interaction, in: HICSS 2020 Proceedings. pp. 5139–5148.

Schoenmueller, V., Netzer, O., Stahl, F., 2018. The Extreme Distribution of Online Reviews: Prevalence, Drivers and Implications. SSRN Electron. J. https://doi.org/10.2139/ssrn.3100217