Comment on “Civic honesty around the globe” – email usage habit predicts how likely to get an email response

Rongfeng Cui^{1*}, Yanhong Hou², Pingze Zhang¹, Wenjie Yao^{3,4,5}, Xinxin Du^{6,7} and Yongwei Guo²

¹Max Planck Institute for Biology of Aging, Cologne, Germany 50931

²Department of Ophthalmology, University of Cologne, Kerpener Str. 62, Cologne Germany 50937

³Institute for Genetics, Centre for Molecular Medicine (CMMC), University of Cologne, Cologne, Germany 50931

⁴Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany 50931

⁵Institute for Neurophysiology, Faculty of Medicine and University Hospital, University of Cologne, Cologne Germany 50931

⁶Department of Gynecology, University of Cologne, Cologne, Germany 50931

⁷Department of Obstetrics and Gynecology, PLA General Hospital, Beijing, China 100853

*Correspondence to: rcui@age.mpg.de

Abstract After correcting for email usage prevalence and other experimental control variables, we found that all country-level correlations originally reported in Cohn et al. vanished. No conclusions can be drawn on cross-country differences in the rates of reporting lost wallets.

Field behavioral studies in humans have been complicated by difficulties in controlling for unobserved variables (1). Although the main aim was to quantify the effect of money incentives on wallet return rate, Cohn et al. (2) also reported a drastic variation in email responding rate among countries, and interpreted them as “objective measurements” of civic honesty. They then claimed that various geographic, cultural, and institutional variables of different countries (2) can explain 65% of the variation in the adjusted email responding rate (their Fig S6).

Email usage reported by companies was cited to correct for the response rates in 27 nations, excluding most western European countries (2). There is, however, no expectation on how email usage by enterprises should correlate with individual telecommunication habits. We instead compiled email usage frequency data among ordinary internet users from publicly available datasets for 36 countries (Table S1). We only analyzed Money and NoMoney conditions given their complete country coverage. The email usage weighted by internet prevalence explained 71% of the variance in observed email response rate in the NoMoney treatment and 67% in the Money treatment (Fig. S1). In addition, we also found response rate to significantly correlate with crime index (Fig. S2), communication accuracy (Fig. S3-5) and population density of sampled cities (Fig. S6).

Principal component (PC) analysis of all independent variables revealed that many country-level variables are strongly loaded on PC1, where weighted email usage has the highest share of loading (Fig. S7), suggesting that country-level variables were confounded by email usage. PC1 alone can explain 75.24% of the variation in response rate. After removing the effects of weighted email usage (Fig. S9), or together with 26 other control variables from the country-level variables by linear regression (Fig. S8), none of their residuals significantly correlate with response rate after false discovery rate correction. We present a corrected Fig. 1 with email response residuals, with the emphasis that the country rankings were unexplained by statistically meaningful factors.

Other caveats that we have not accounted for here includes pseudoreplication of countries of shared cultural history, akin to the non-independency problem in phylogenetic comparisons (3), which results in over-confidence of significant tests, if any.

Our analyses confirm that money incentives increase email response rate, but cross-country differences were mainly driven by spurious correlations unrelated to civic honesty.

References

1. J. Henrich, S. J. Heine, A. J. N. Norenzayan, Most people are not WEIRD. *Nature* **466**, 29 (2010).
2. A. Cohn, M. A. Maréchal, D. Tannenbaum, C. L. Zünd, Civic honesty around the globe. *Science*, eaau8712 (2019).
3. J. Felsenstein, Phylogenies and the comparative method. **125**, (1985).

Figures

Fig. 1. Residuals of email response rates of 36 countries, correcting for 25 variables related to experimental conditions, population density, communication accuracy and email usage habits (Table S2). Yellow dot – NoMoney treatment. Red dot – Money treatment. Variations between countries were not explained by any examined country-level variables (Fig. S6, Fig. S7).

Supplementary Information

Compilation of Email and internet usage data

We searched internet databases and webpages for email usage frequencies, resulted from surveys of internet users or compiled by government agencies. We prioritize on statistics reported for 2015, and used the closest date to 2015 if that year is not available. If multiple statistics are stratified by, for example, age or sex, we took the average number. Together we obtained email usage statistics for 36 countries (Table S1; Supplementary data).

We downloaded global internet prevalence data from the International Telecommunication Union website for the year 2015. In cases where email usage on computer and mobile devices were available, we took the computer estimate. In datasets with indication of how often users email, we summed the proportion of users that used emails at least once per month. We emphasize that while the absolute proportions may be higher in the sampled cities due to the fact that they are the most developed part of the countries, these numbers should provide a good approximation on the relative tendency for locals to use email communication. The weighted email usage is computed by multiplying the email usage frequency by the internet prevalence in each country.

Approximating communication accuracy

A commonly encountered issue of cross-cultural studies is the language barrier, among other cultural differences. English was chosen as the language for instruction in this study, regardless of countries and their native language. When the test subject failed to understand English, the research assistant depended on a translation app to communicate. In this dataset for example, Chinese subjects were subjectively scored by the research assistants to have a level of understanding matching the Germans, with completely overlapping distributions. Counterintuitively, while >90% of the German subjects understood English, the research assistants used a translation app to talk to more than 50% of the Chinese subjects (Fig. S4A). A similar pattern (Fig. S4B) was found between Indonesia and Switzerland (among other European countries). Unless one assumes the translator app performed equally well as human, the explanation can only be that research assistants were overconfident on the level of understanding by the subjects in a different culture. We examined translation results of three commonly used apps on the phrase “Can you please take care of it”. Two (google translate, Microsoft Bing) out of three translators gave the translation in Chinese equivalent to “Can you look after it”, meaning that the subject should only keep the wallet safe but not take other actions. Translation errors or ambiguities may explain why subjects did not actively seek owners.

To statistically account for different accuracies in translation and difficulties in communication in English, we computed a composite score as an approximation. First, we use a previously published language translator accuracy matrix for Google Translate (1), and averaged the accuracies (standardized to the range 0-1) for translating to and from English. For test subjects who communicated directly using English during the interaction with researchers, we also account for their English ability by using nation-wide TOEFL scores standardized to the range 0-1, published by ETS. Countries with English as the official language are coded with a standardized TOEFL score of 1. The final composite score, which we call “Communication Accuracy”, is computed as $\text{EnglishSpeaker} * \text{TOEFL} + \text{NonEnglishSpeaker} * \text{TranslatorAccuracy}$.

We found that in the deposited data file, “no_english” was incorrectly coded as “1” for India and Kenya, whose official language included English. We corrected them to “0” for subsequent analyses, following Cohn et al.’s definition.

Communication accuracy was positively correlated with response rate, explaining 15.9% of the variance in the NoMoney treatment and 21.0% in the Money treatment (Fig. S5; Fig. S7).

Crime index

One possible explanation for subjects to not respond is that they have more experience with scams, typical of countries with more crimes, making it more difficult to persuade subjects of the genuineness of the research assistant. This hypothesis is not far-fetched, because one research assistant was arrested by the police in Kenya due to suspicious activities. We downloaded Crime index in 2015 from https://www.numbeo.com/crime/rankings_by_country.jsp?title=2015 for correlations. As expected, the crime index is negatively correlated with response rate (Fig. S2; Fig. S7). We note that this variable was not considered as a control variable.

Population density

Cohn et al. observed how busy the subject was for a short period. This should have been an important control variable, because one expects that a more busy subject would be less likely to return an email. It is surprising to find, however, that busy rating was instead *positively* correlated with email response rate (Fig. S7).

We instead hypothesize that how busy a person is should correlate with local population density, and may influence the response rate because employees would have less time dealing with non-business requests.

We compiled population density data from the English version of Wikipedia, which in turn cited data from respective governmental sources. When data were available for different areas of a metropolitan area, only the density in the more central urban region was used because there is a large variation on the suburban regions different cities administer. When we cannot determine whether the data referred to the urban or the greater metropolitan area, we excluded them. Because the raw dataset in excel format deposited by Cohn et al. contained only numeric city codes, without referring to the actual city names, we were unable to associate the city names in the supplementary information to the excel table. However, because the sampling appeared to be quite even between different cities within a country, we simply averaged the population densities for all sampled cities in a country.

We indeed see a negative correlation with log population density with email response rate (Fig. S6), as expected if it is a proxy for how busy people are. We consider it to be a control variable.

Principal component analysis

Due to the very high number of independent variables (including control variables and candidate country-level variables) in the dataset, it is necessary to remove their self-correlations prior to subsequent analysis, to avoid problems caused by multicollinearity during linear regressions (2). Due to extensive missing data, the variable “r_PL_propertylaw” was excluded from the matrix. The following variables contained sporadic missing data: “security_cam”, “security_guard”, “c_WVS_familyties”, “c_GPLS_judicial_independence”, “c_GPLS_constitutional_review”, “c_GPLS_plurality” and “c_GPLS_prop_representation”. In order to maximize usable data, we imputed the missing values with the missingMDA R package, using all other independent variables (but not response rate) as predictors.

The optimal number of PCs were then estimated on the matrix with missing data using the *estim_ncpPCA* R function using the kfold approach, with a maximal number of PCs set to 30. This procedure identified 18 PCs to be sufficient to describe all independent variables. The PCA was performed using the data matrix with imputed cells, with the component number set to 18. We then used a model selection approach to identify which PCs significantly correlated with email response rates.

For model selection, we use a stepwise variable addition/elimination approach, implemented in the R `stepAIC()` function, using the Akaike information criterion (AIC) during each iteration to determine the most parsimonious model that yields the lowest AIC. The addition/elimination steps were run in both the direction of increasing and decreasing model complexity (`direction="both"`), starting from the simplest model including only the intercept. Only main effects were examined. We note that the model choice was robust to the choice of the starting model and the direction of iteration.

This procedure yielded an optimal model containing 14 out of the 18 PCs, explaining 89.35% of the variation in email response rate. Among these, PC1 explains 75.24% while PC12 explains 3.83%. The rest of the PCs only explain a small proportion of the variation. We visualize the variable loadings onto each PC, and found that money condition uniquely loaded onto PC12, suggesting that this variable is unlikely to be confounded by other effects (Fig. S6). However, weighted email usage most heavily loaded onto PC1, where many other country-level variables also have significant loadings, suggesting that their correlations were confounded by email usage (Fig. S6).

We next removed the influence of situational and experimental variables (control variables, bolded in Fig. S6) from the candidate country-level variables. First, we performed PCA on all control variables and correlated all PCs with email response rate. Significant PCs ($p < 0.05$) were used to perform linear regressions with the country-level variables. By taking the residuals from these regressions, we effectively remove any spurious correlations caused by weighted email usage and other idiosyncratic experimental conditions. None of the residuals correlated with email response rate after FDR correction (Fig. S7) any longer.

Even when we only corrected for weighted email usage, no country-level variables significantly correlated with email response rate after FDR correction (Fig. S8).

We produced a new Fig. 1 showing the residual email response rates corrected for all control variables, using the same methods outlined above. A prominent feature of this figure is a random distribution of developed and developing nations and a narrower spread of data.

Supplementary Tables

Table S1. Email usage and other data compiled from internet databases. For population density, the statistics from all visited cities were averaged for the country. Other statistics were directly looked up at the country level. Whenever possible, data from the year 2015 were used.

Country	Email Usage %	Internet Prevalence %	Weighted Email Usage	PopDensity/km ²	NativeLang	TOEFL	Translator Accuracy	CrimeIndex
Kenya	30.0	16.6	5.0	5018.25	English	82	1.00	72.61
Indonesia	25.4	22.0	5.6	7301.41	Indonesian	85	0.94	47.22
Morocco	19.0	57.1	10.8	6582.91	Arabic	81	0.63	47.55
Turkey	25.0	53.7	13.4	1710.03	Turkish	78	0.41	36.40
China	28.4	50.3	14.3	4332.72	Chinese	79	0.33	41.75
South Africa	33.0	51.9	17.1	1667.20	Afrikaans	98	0.48	78.44
Malaysia	25.0	71.1	17.8	3140.48	Malay	91	0.42	69.97
Peru	45.9	40.9	18.8	4297.18	Spanish	84	0.67	57.63
India	80.0	26.0	20.8	11333.11	English	94	1.00	46.55
Thailand	53.7	39.3	21.1	4303.18	Thai	78	0.47	37.06
Ghana	71.0	31.5	22.3	17949.24	English	81	1.00	49.42
Brazil	40.0	58.3	23.3	5533.01	Portuguese	87	0.71	68.95
Romania	43.0	55.8	24.0	2834.40	Romanian	91	0.62	29.72
Portugal	44.0	68.6	30.2	2387.72	Portuguese	94	0.71	35.84
Italy	59.0	58.1	34.3	3175.83	Italian	91	0.88	47.50
Greece	53.0	66.8	35.4	6305.70	Greek	93	0.80	41.35
Croatia	53.0	69.8	37.0	2014.82	Croatian	90	0.46	29.34
Poland	58.0	68.0	39.4	1998.21	Polish	90	0.64	32.99
Argentina	61.5	68.0	41.8	5670.00	Spanish	90	0.67	62.40
Spain	54.0	78.7	42.5	3940.81	Spanish	89	0.67	33.85
Mexico	77.0	57.4	44.2	4386.91	Spanish	86	0.67	53.84
Israel	62.0	77.4	48.0	5693.71	Hebrew	91	0.35	30.40
Chile	68.0	76.6	52.1	2513.48	Spanish	81	0.67	45.01
France	67.0	78.0	52.3	4738.73	French	88	0.92	49.71
Canada	64.0	88.5	56.6	2276.39	French	95	0.92	37.46
Czech Republic	79.0	75.7	59.8	1475.26	Czech	89	0.57	32.89
Germany	73.0	84.4	61.6	2958.60	German	98	0.82	28.49
USA	90.8	74.6	67.7	2212.95	English	89	1.00	50.01
UK	76.0	92.0	69.9	3425.65	English	96	1.00	42.01
Australia	84.0	84.6	71.0	2202.97	English	92	1.00	42.16
Sweden	90.0	90.6	81.5	2964.75	Swedish	93	0.89	42.09
Netherlands	91.0	91.7	83.5	3624.50	Dutch	100	0.82	35.41
Denmark	88.0	96.3	84.8	1776.01	Danish	98	0.89	25.71
Switzerland	97.0	87.5	84.9	3418.96	German	99	0.82	26.77
New Zealand	99.0	88.2	87.3	927.40	English	95	1.00	39.76
Norway	91.0	96.8	88.1	1694.64	Norwegian	93	0.79	31.19
Kazakhstan	NA	70.8	NA	1345.78	Russian	84	0.78	46.68
Russia	NA	NA	NA	3111.48	Russian	87	0.78	51.33
Serbia	NA	65.3	NA	1699.96	Croatian	87	0.46	43.43
UAE	NA	90.5	NA	2276.75	Arabic	84	0.63	22.22

Supplementary Figures

Fig. S1. Strong correlation between weighted email usages in 36 countries with the email response rate. Weighted email usage is computed by multiplying the frequency of internet usage in each country in 2015 with the surveyed proportion of internet users that use email for communication.

Fig. S2. Strong negative correlation between crime index and email response rate.

Baidu

Google

Microsoft Bing

Fig. S3. Screenshots of results from English to simplified Chinese using 3 popular mobile translation apps. Taken in June, 2019. Two (Google and Microsoft Bing) out of three translators mistranslated the instruction “take care of it” into Chinese “look after it”.

Fig. S4. The distribution of the subjective assessment of whether the test subject understood the situation did not differ between (A) Germany and China ($p=0.147$, Wilcoxon rank sum test) and (B) Switzerland and Indonesia ($p=0.273$, Wilcoxon rank sum test), yet significantly more Germans and Swiss were able to speak English with the research assistants ($p=4.36e-15$ and $p=3.48e-61$, Wilcoxon rank sum test). Suggesting potential overconfidence in the researcher's assessment of how well the test subjects understood the instructions.

Fig. S5. Correlation between communication accuracy and email response rate. Communication accuracy = (english_speaking * TOEFL) + (non_english_speaking * TranslatorAppAccuracy).

Fig. S6. Correlation between log10(average population density) and email response rate. Population density is averaged among all visited cities in a country. Only urban regions were included in the calculation.

Fig S7. Visualization of principal component loadings of all independent variables, including experimental control (in bold) and candidate country-level explanatory variables (in normal font). The signs of the loadings were calculated such that it is in the direction of predicting higher email response rates. The relative loadings were multiplied by the percentage of variation explained by the PC shown on row 1. PC1 explains the majority of the variation, and receives the heaviest loading from the variables weighted email usage. This suggests that other factors are probably correlated with email response rate through their correlation with the weighted email usage.

Fig. S8. After removing the correlation with all potential confounding experimental variables (including "money_condition", "weighted_email_usage", "email_usage", "internet_usage", "computer", "TOEFL", "TranslatorAccuracy", "no_english", "security_cam", "cultural", "CommunicationAccuracy", "busy", "understood_situation", "rejection_frac", "hotel", "other_bystanders", "postal", "public", "age", "maleexperimenter", "bank", "age_experimenter", "logPopDensity", "male", "security_guard" and "coworkers"), none of the other variables were significantly correlated with email response ($FDR > 0.2$). Bar plot shows the correlation coefficients, error bars are standard errors of the means.

Fig. S9. After removing the correlation with weighted email usage and money treatment condition, none of the other variables were significantly correlated with email response (FDR > 0.2). Bar plot shows the correlation coefficients, error bars are standard errors of the means.

References

1. M. Aiken, S. Balan, An analysis of Google Translate accuracy. *Translation Journal* **16**, 1-3 (2011).
2. M. H. Graham, Confronting multicollinearity in ecological multiple regression. *Ecology* **84**, 2809-2815 (2003).
3. T. Yamashita, K. Yamashita, R. J. C. i. S. T. Kamimura, Methods, A stepwise AIC method for variable selection in linear regression. *Communications in Statistics—Theory and Methods* **36**, 2395-2403 (2007).