# Online Appendix

Figure A.1 Screenshot of intervention pop-up message

Place “Fig\_A1.tiff” here

Notes: Public Benefit + Privacy treatment displayed.

Figure A.2 Increase in number of old entries filled by treatment

Place “Fig\_A2.tiff” here

Notes: The figure displays the post-intervention increase in the number of profile entries completed. The error bars indicate the 95% confidence intervals. Old entries exclude most recently added profile fields “main motivation” and “regular computer use.”

Table A1 OLS regression results of further outcomes with controls by treatment

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Sensitive entries |  | Type of changes |  |  |  |  |  |  |
|  | Yes | No | Extensions | Deletions | Updates |  |  |  |  |
|  | (1) | (2) | (3) | (4) | (5) |  |  |  |  |
| Benefit | 0.060\*\* | 0.073\* | 0.186\*\* | 0.000 | 0.008 |  |  |  |  |
|  | (0.027) | (0.038) | (0.080) | (0.003) | (0.009) |  |  |  |  |
| Public Benefit + Privacy | 0.072\*\*\* | 0.070\* | 0.226\*\*\* | 0.002 | 0.012 |  |  |  |  |
|  | (0.028) | (0.038) | (0.081) | (0.003) | (0.009) |  |  |  |  |
| Entries pre | 0.853\*\*\* | 1.026\*\*\* | -0.118\*\*\* | 0.003\*\*\* | 0.019\*\*\* |  |  |  |  |
|  | (0.006) | (0.004) | (0.008) | (0.001) | (0.002) |  |  |  |  |
| Constant | 1.103\*\*\* | 1.804\*\*\* | 1.030\*\*\* | -0.000 | 0.001 |  |  |  |  |
|  | (0.047) | (0.060) | (0.134) | (0.005) | (0.012) |  |  |  |  |
| N | 6155 | 6155 | 6155 | 6155 | 6155 |  |  |  |  |
| R2 | 0.64 | 0.78 | 0.03 | 0.01 | 0.05 |  |  |  |  |

*Notes:* The table reports OLS regression results on the intensive margin. Robust standard errors in parentheses. All specifications estimated with controls including dummies for the courses “International Teams,” “50 Years of Internet,” and “Data Science & Engineering,” whether the course is the first course on the platform, whether the course is accessed from Germany, and whether it is accessed earlier than the median access, as well as the day of enrollment relative to the course start. “Entries pre” for sensitive and insensitive categories corresponds to only those ex ante filled entries classified as sensitive or insensitive, respectively. All “Entries pre” are transformed to a mean of zero. This way the constant can be interpreted as the effect observed in the *Control* group. \* *p*<0.10, \*\* *p*<0.05, \*\*\* *p*<0.01.

Table A2 OLS regression results of disclosing any (in)sensitive entry on treatment

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sensitive |  | Insensitive |  |
|  | (1) | (2) | (3) | (4) |
| Benefit | 0.023\*\* | 0.023\*\* | -0.007 | -0.007 |
|  | (0.010) | (0.010) | (0.006) | (0.006) |
| Public Benefit + Privacy | 0.023\*\* | 0.023\*\* | -0.007 | -0.007 |
|  | (0.010) | (0.010) | (0.006) | (0.006) |
| At least one entry | 0.795\*\*\* | 0.794\*\*\* | 0.918\*\*\* | 0.919\*\*\* |
| (pre-intervention) | (0.006) | (0.007) | (0.005) | (0.006) |
| Constant | 0.190\*\*\* | 0.157\*\*\* | 0.087\*\*\* | 0.079\*\*\* |
|  | (0.008) | (0.017) | (0.007) | (0.010) |
| Controls | No | Yes | No | Yes |
| N | 6155 | 6155 | 6155 | 6155 |
| R2 | 0.58 | 0.58 | 0.87 | 0.87 |

 *Notes:* The table reports OLS regression results on the extensive margin. Robust standard errors in parentheses. Controls include dummies for the courses “International Teams,” “50 Years of Internet,” and “Data Science & Engineering,” whether the course is the first course on the platform, whether the course is accessed from Germany, and whether it is accessed earlier than the median access, as well as the day of enrollment relative to the course start. “At least one entry (pre-intervention)” for sensitive and insensitive categories corresponds to only those ex ante filled entries classified as sensitive or insensitive, respectively. \* *p*<0.10, \*\* *p*<0.05, \*\*\* *p*<0.01.

Heterogeneity on the intensive margin

Besides our main hypotheses, we check for heterogeneity between different groups of users defined by always available platform process data. First, treatment effects may be stronger for first-time users, i.e., user for whom the treated course is the first course they take on the MOOC platform. These users have no experience with the platform so they more likely underestimate the size of the user community and hence the public benefits of personal data contribution compared to users with previous course experience. Hence, first-time users may show a stronger reaction to the *Public Benefit* treatment. However, we find no support for such an effect. In an OLS regression reported in column (1) of Table A.3 the interaction effect between first-time course taking and *Public Benefit* is insignificant and small in magnitude.

Besides the potentially underestimated public benefit, less experience with the platform may also mean that first-time users have less trust in the platform so they may be less willing to share sensitive personal data with the platform than experienced users. Mitigating this obfuscation may therefore additionally lead to stronger responses to the *Public Benefit + Privacy* treatment for first-time compared to experienced users. While there is a sizeable interaction term of *Public Benefit + Privacy* with first-time course taking, it is not distinguishable from the interaction with *Public Benefit* at conventional significance levels ($p=0.193$). Consequently, we find no clear support for first time users reacting differently to any of our treatments than experienced users.

Second, we divide our sample based on pre-intervention commitment to the course, i.e., we call a user committed if she shows up in the course earlier than the median user to work on the first week’s course material. We expect that committed users are more willing to reciprocate the platform’s course offer by contribution data. Yet, we do not find support for such an effect. As column (2) of Table A.3 reports, the interactions of early first course action with *Public Benefit* and *Public Benefit + Privacy* are negative, small in magnitude, and not statistically significant at any conventional level. Hence, commitment to the platform in form of early course action does not lead to more reciprocal behavior via contributing more personal data.

Third, heterogeneous responses to treatments may exists between Germans and non-Germans. Since previous research indicates that Germans hold comparably strong privacy concerns (IBM 2018; Bellman et al. 2004), the *Public Benefit + Privacy* treatment may generate a stronger positive effect on personal information disclosure for Germans than non-Germans. We measure this variable based on whether the most frequently used platform access location when logged in lies in Germany or not. However, we do not find that the privacy protection emphasis in *Public Benefit + Privacy* makes German users more willing to contribute their data. While the interaction effect of the *Public Benefit + Privacy* dummy with the Germany dummy in column (3) of Table A.3 is larger than zero, it is not significant. Moreover, it is smaller in magnitude than the interaction effect with *Public Benefit* and statistically indistinguishable in a joint test for equal reactions across treatments as specified in the last row ($p=0.720$). Thus, in our context, there is no statistical support for Germans reacting more to privacy protection than non-Germans.

Fourth, users with few profile fields filled before the intervention may be more prone to privacy concerns regarding data sharing. In order to study this, we split our sample into three categories: user with no initially filled profile entries, users with at least one but not more than median entries, and users with more than median entries. If privacy concerns limit initial disclosure, we may see more additional entries generated in the *Public Benefit + Privacy* treatment for users with relatively few initial entries due to salient privacy protection standards in this treatment. Nonetheless, we do not find evidence that users with different many ex ante entries react differently to the intervention. There are no significant interaction effects of the pre-intervention profile completion status and the *Public Benefit* and *Public Benefit + Privacy* treatment dummies in column (4) of Table A.3, and the reactions of users with zero or many initial entries, respectively, to *Public Benefit* and *Public Benefit + Privacy* are indistinguishable ($p=0.719$ and $p=0.860$, respectively). Consequently, emphasizing privacy protection in the *Public Benefit + Privacy* treatment does not motivate user with fewer initial entries to disclose more.

Table A3 OLS regression results: Heterogeneity by treatment

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | First course | First action | Germany | Entries pre |
|   | (1) | (2) | (3) | (4) |
| Benefit  |  0.184\*\*  | 0.195  |  0.005  |  0.064 |
|   |  (0.085)  |  (0.124)  |  (0.237)  |  (0.164) |
| Public Benefit + Privacy  |  0.161\*  |  0.249\*\* |  0.085  |  0.069 |
|   |  (0.085)  |  (0.127)  |  (0.233)  |  (0.167)  |
| Entries pre  |  0.879\*\*\* |  0.879\*\*\*  |  0.879\*\*\*  |  |
|   |  (0.008)  |  (0.008)  |  (0.008) |  |
| First Course  |  -0.191  |  |  |  |
|   |  (0.164)  |  |  |  |
| Benefit x First Course  |  0.014  |   |   |   |
|  |  (0.232)  |   |   |   |
| Public Benefit + Privacy x First Course  |  0.328  |   |   |   |
|   |  (0.238)  |   |   |   |
| Early First Action  |  |  0.123  |  |  |
|   |  |  (0.111)  |  |  |
| Benefit x Early First Action  |   |  -0.015  |   |   |
|  |   |  (0.162)  |   |   |
| Public Benefit + Privacy x Early First Action  |   |  -0.041  |   |   |
|  |   |  (0.165)  |   |   |
| Germany  |  |  |  -0.147  |  |
|  |   |  |  (0.176)  |  |
| Benefit x Germany  |   |   |  0.208 |   |
|  |   |   |  (0.252)  |   |
| Public Benefit + Privacy x Germany  |   |   |  0.159  |   |
|  |   |   |  (0.248)  |   |
| Zero Entries Pre  |   |   |   |  -3.246\*\*\*  |
|  |   |   |   |  (0.149)  |
| Many Entries Pre  |   |   |   |  3.842\*\*\*  |
|  |   |   |   |  (0.133)  |
| Benefit x Zero Entries Pre  |   |   |   |  0.270  |
|  |   |   |   |  (0.309)  |
| Benefit x Many Entries Pre  |  |  |  |  -0.082  |
|  |   |   |   |  (0.267)  |
| Public Benefit + Privacy x Zero Entries Pre  |   |   |   |  0.164  |
|  |   |   |   |  (0.212)  |
| Public Benefit + Privacy x Many Entries Pre  |   |   |   |  0.122  |
|  |   |   |   |  (0.190)  |
| Constant |  1.302\*\*\* |  1.271\*\*\*  |  1.392\*\*\* |  4.434\*\*\* |
|   | (0.130)  |  (0.136)  |  (0.180)  |  (0.168) |
| N  | 6155 | 6155 | 6155 | 6155 |
| R2  |  0.55  |  0.55  |  0.55  |  0.48  |
|  |  |  |  |  |
| p-value: Benefit + Benefit x Subgroup =  | 0.193 | 0.799 | 0.720 | 0.719 |
|  Benefit+Cost + Benefit+Cost x Subgroup  |  |  |  |  0.860 |

  *Notes:* Table reports OLS regression results on the intensive margin. Robust standard errors in parentheses. All specifications estimated with controls including dummies for the courses “International Teams,” “50 Years of Internet,” and “Data Science & Engineering.” Additional controls for first course, early first course action, and access from Germany are always included even if not listed in the table. The last row reports p-values of testing for treatment differences between *Public Benefit* and *Public Benefit + Privacy* for the respective user subgroups in the column. *p*<0.10, \*\* *p*<0.05, \*\*\* *p*<0.01.

In sum, we find no evidence for heterogeneous responses between treatments for subgroups defined by process data. In other words, any version of the pop-up prompting users to complete their profile will attract information from all subgroups alike. This suggests choosing the most effective version in terms of intensive margin changes, the *Public Benefit + Privacy* message, to provide the platform with a solid non-selective boost in available information.

Figure A.3 Histograms of profile entry content before and after the intervention

Place here

1. Position

“Fig\_A3.A.tiff”

1. Career Status

“Fig\_A3.B.tiff”

1. Professional Experience

“Fig\_A3.C.tiff”

1. Highest degree

“Fig\_A3.D.tiff”

1. Gender

“Fig\_A3.E.tiff”

1. Age group

“Fig\_A3.F.tiff”

1. Motivation

“Fig\_A3.G.tiff”

1. Regular computer use

“Fig\_A3.H.tiff”

Notes: Figure displays histograms of which data content is contributed

Table A4 Content of pre-intervention profile entries

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Control |  | Public Benefit |  | Public Benefit + Privacy |  |  |
| Outcome | Category | Mean | Std dev | Mean | Std dev | Mean | Std dev |  |
| position | *missing* | 0.777 | 0.416 | 0.789 | 0.408 | 0.798 | 0.401 |  |
|  | department head | 0.035 | 0.184 | 0.030 | 0.170 | 0.023 | 0.152 |  |
|  | intern | 0.005 | 0.070 | 0.011 | 0.105 | 0.008 | 0.091 |  |
|  | project manager | 0.068 | 0.251 | 0.068 | 0.253 | 0.070 | 0.254 |  |
|  | team leader | 0.032 | 0.176 | 0.025 | 0.155 | 0.029 | 0.168 |  |
|  | technician | 0.083 | 0.276 | 0.077 | 0.267 | 0.071 | 0.258 |  |
| affiliation | *missing* | 0.887 | 0.316 | 0.906 | 0.291 | 0.893 | 0.309 |  |
| career status | *missing* | 0.667 | 0.472 | 0.669 | 0.471 | 0.676 | 0.468 |  |
|  | academic researcher | 0.011 | 0.105 | 0.010 | 0.100 | 0.009 | 0.093 |  |
|  | other | 0.048 | 0.214 | 0.038 | 0.192 | 0.046 | 0.210 |  |
|  | professional | 0.243 | 0.429 | 0.244 | 0.429 | 0.229 | 0.420 |  |
|  | student | 0.015 | 0.120 | 0.021 | 0.143 | 0.018 | 0.132 |  |
|  | teacher | 0.017 | 0.128 | 0.018 | 0.133 | 0.022 | 0.147 |  |
| professional  | *missing* | 0.680 | 0.467 | 0.684 | 0.465 | 0.691 | 0.462 |  |
| life | more than 10 years | 0.243 | 0.429 | 0.245 | 0.430 | 0.234 | 0.423 |  |
|  | up to 10 years | 0.041 | 0.198 | 0.036 | 0.187 | 0.037 | 0.188 |  |
|  | up to 5 years | 0.036 | 0.186 | 0.034 | 0.181 | 0.039 | 0.193 |  |
| highest degree | *missing* | 0.671 | 0.470 | 0.673 | 0.469 | 0.678 | 0.467 |  |
|  | bachelor | 0.031 | 0.174 | 0.033 | 0.179 | 0.034 | 0.182 |  |
|  | diplom | 0.092 | 0.289 | 0.097 | 0.296 | 0.088 | 0.283 |  |
|  | high school student | 0.057 | 0.231 | 0.053 | 0.225 | 0.050 | 0.219 |  |
|  | magister | 0.006 | 0.076 | 0.006 | 0.079 | 0.008 | 0.088 |  |
|  | master | 0.058 | 0.233 | 0.057 | 0.232 | 0.059 | 0.236 |  |
|  | other | 0.049 | 0.216 | 0.054 | 0.227 | 0.059 | 0.236 |  |
|  | phd | 0.037 | 0.188 | 0.026 | 0.158 | 0.023 | 0.150 |  |
| gender | *missing* | 0.642 | 0.479 | 0.659 | 0.474 | 0.646 | 0.478 |  |
|  | female | 0.051 | 0.220 | 0.055 | 0.229 | 0.064 | 0.244 |  |
|  | male | 0.307 | 0.461 | 0.285 | 0.452 | 0.291 | 0.454 |  |
| age group | *missing* | 0.604 | 0.489 | 0.620 | 0.485 | 0.611 | 0.488 |  |
|  | 20-29 | 0.023 | 0.150 | 0.026 | 0.158 | 0.027 | 0.162 |  |
|  | 30-39 | 0.079 | 0.270 | 0.074 | 0.262 | 0.072 | 0.258 |  |
|  | 40-49 | 0.093 | 0.291 | 0.096 | 0.294 | 0.097 | 0.296 |  |
|  | 50-59 | 0.121 | 0.326 | 0.115 | 0.319 | 0.114 | 0.318 |  |
|  | 60-69 | 0.048 | 0.214 | 0.042 | 0.201 | 0.054 | 0.226 |  |
|  | 70+ | 0.027 | 0.162 | 0.018 | 0.135 | 0.017 | 0.130 |  |
|  | <20 | 0.004 | 0.066 | 0.009 | 0.093 | 0.008 | 0.088 |  |
| motivation | *missing* | 0.947 | 0.224 | 0.954 | 0.210 | 0.945 | 0.228 |  |
|  | credits | 0.000 | 0.022 | 0.001 | 0.031 | 0.000 | 0.000 |  |
|  | other | 0.000 | 0.000 | 0.000 | 0.022 | 0.000 | 0.000 |  |
|  | personal | 0.016 | 0.126 | 0.016 | 0.124 | 0.013 | 0.114 |  |
|  | professional | 0.037 | 0.188 | 0.029 | 0.168 | 0.042 | 0.200 |  |
| computer use | *missing* | 0.947 | 0.224 | 0.950 | 0.219 | 0.943 | 0.231 |  |
|  | easy | 0.001 | 0.031 | 0.003 | 0.058 | 0.001 | 0.038 |  |
|  | high | 0.028 | 0.164 | 0.026 | 0.160 | 0.031 | 0.173 |  |
|  | intermediate | 0.024 | 0.154 | 0.021 | 0.143 | 0.024 | 0.155 |  |

 *Notes:* Table reports mean shares of entry content pre-intervention and corresponding standard deviations.

Table A5 Marginal effects of multinominal logit regressions regarding pre-post shifts in profile content distributions

|  |  |  |  |
| --- | --- | --- | --- |
| Outcome | Category | Marginal effect | Standard error |
| position | department head | 0.0173\*\*\* | 0.006 |
|  | intern | -0.0035 | 0.003 |
|  | project manager | 0.0068 | 0.008 |
|  | team leader | -0.0022 | 0.006 |
|  | technician | -0.0183\*\* | 0.007 |
| career status | academic researcher | 0.0031 | 0.003 |
|  | other | 0.0031 | 0.005 |
|  | professional | 0.0007 | 0.006 |
|  | student | -0.0016 | 0.003 |
|  | teacher | -0.0054\*\* | 0.002 |
| professional life | more than 10 years | 0.0116\* | 0.006 |
|  | up to 10 years | -0.0123\*\* | 0.005 |
|  | up to 5 years | 0.0007 | 0.005 |
| highest degree | high school student | -0.0107\*\* | 0.005 |
|  | bachelor | 0.0013 | 0.004 |
|  | master | 0.0397\*\*\* | 0.007 |
|  | phd | 0.0047 | 0.004 |
|  | other | -0.0350\*\*\* | 0.005 |
| gender | female | 0.0176\*\*\* | 0.005 |
|  | male | -0.0176\*\*\* | 0.005 |
| age group | 20-29 | 0.0123\*\*\* | 0.003 |
|  | 30-39 | 0.0046 | 0.004 |
|  | 40-49 | -0.0072\*\* | 0.004 |
|  | 50-59 | -0.0085\*\* | 0.004 |
|  | 60-69 | -0.0043\* | 0.003 |
|  | 70+ | -0.0004 | 0.002 |
|  | <20 | 0.0035\*\* | 0.002 |
| motivation | credits | -0.0007 | 0.005 |
|  | other | 0.0124 | 0.010 |
|  | personal | 0.0305 | 0.024 |
|  | professional | -0.0421\* | 0.025 |
| computer use | easy | 0.0229\* | 0.014 |
|  | high | -0.0319 | 0.025 |
|  | intermediate | 0.0091 | 0.025 |

 *Notes:* Table reports average marginal effects from multinominal logit regressions for each outcome in the leftmost column. We combined the more traditional degrees “Magister” and “Diplom” with “Master” because they belong to the same International Standard Classification of Education level. \* *p*<0.10, \*\* *p*<0.05, \*\*\* *p*<0.01.