APPENDIX: FOR ONLINE PUBLICATION ONLY

A Appendix Tables

Table A.1: Attrition, Team Level

	Non-Attrited (1)	Attrited (2)	Diff. (3)	Std. Diff. (4)
Gender-mixed team	0.330	0.362	0.032	0.047
	(0.471)	(0.484)	(0.062)	
All-female team	0.336	0.304	-0.032	-0.048
	(0.473)	(0.464)	(0.062)	
Mean A-level GPA	2.741	2.724	-0.017	-0.074
	(0.165)	(0.150)	(0.021)	
Share top-tier high school	0.828	0.815	-0.013	-0.048
	(0.191)	(0.190)	(0.025)	
Mean age	22.687	22.525	-0.162	-0.080
	(1.500)	(1.369)	(0.195)	
Share foreign nationality	0.036	0.065	0.029	0.205
	(0.092)	(0.111)	(0.013)	
Share study program Master level	0.243	0.214	-0.030	-0.098
	(0.203)	(0.224)	(0.027)	
Share study program arts and humanities	0.241	0.283	0.041	0.134
	(0.210)	(0.227)	(0.028)	
Share study program engineering	0.192	0.188	-0.004	-0.013
	(0.214)	(0.194)	(0.028)	
Share study program natural sciences	0.102	0.069	-0.033	-0.181
	(0.147)	(0.112)	(0.019)	
Share study program economics and business	0.289	0.272	-0.018	-0.054
	(0.240)	(0.222)	(0.031)	
N. of obs.	342	69	411	411

Notes: This table documents attrition at team level. Attrition happens because teams are disqualified if a member drops out during the team task. Column (1) shows means and standard deviation for non-attrited teams. Column (2) shows means and standard deviation for attrited teams. Column (3) shows estimated differences between attrited and non-attrited teams and corresponding standard errors. Column (4) shows standardized differences.

Table A.2: Attrition, Individual Level (First Stage)

	Non-Attrited	Attrited	Diff.	Std. Diff.
	(1)	(2)	(3)	(4)
Gender-mixed team	0.330	0.362	0.032	0.047
	(0.471)	(0.482)	(0.063)	
All-female team	0.336	0.304	-0.032	-0.048
	(0.473)	(0.461)	(0.061)	
A-level GPA	2.741	2.724	-0.017	-0.019
	(0.613)	(0.635)	(0.020)	
Top-tier high school	0.828	0.815	-0.013	-0.024
	(0.377)	(0.389)	(0.025)	
Age	22.687	22.525	-0.162	-0.038
	(3.143)	(2.890)	(0.183)	
Foreign nationality	0.036	0.065	0.029	0.095
	(0.186)	(0.247)	(0.014)	
Study program: Master level	0.243	0.214	-0.030	-0.050
	(0.429)	(0.411)	(0.029)	
Study program: Arts and humanities	0.241	0.283	0.041	0.067
	(0.428)	(0.451)	(0.029)	
Study program: Engineering	0.192	0.188	-0.004	-0.007
	(0.394)	(0.392)	(0.026)	
Study program: Natural sciences	0.102	0.069	-0.033	-0.085
	(0.303)	(0.254)	(0.016)	
Study program: Economics and business	0.289	0.272	-0.018	-0.028
	(0.454)	(0.446)	(0.030)	
N. of obs.	1368	276	1644	1644

Notes: This table documents attrition at individual level in the first stage of the experiment. Attrition happens because all members of a team are disqualified if a member drops out during the team task. Column (1) shows means and standard deviation for non-attrited individuals. Column (2) shows means and standard deviation for attrited individuals. Column (3) shows estimated differences between attrited and non-attrited individuals and corresponding standard errors. Column (4) shows standardized differences.

Table A.3: Attrition, Individual Level (Second Stage)

	Non-Attrited	Attrited	Diff.	Std. Diff.
	(1)	(2)	(3)	(4)
Gender-mixed team	0.326	0.393	0.067	0.099
	(0.469)	(0.489)	(0.045)	
All-female team	0.317	0.332	0.015	0.022
	(0.466)	(0.472)	(0.042)	
A-level GPA	2.740	2.737	-0.004	-0.004
	(0.615)	(0.618)	(0.043)	
Top-tier high school	0.818	0.843	0.025	0.047
	(0.386)	(0.365)	(0.027)	
Age	22.648	22.991	0.343	0.074
	(3.052)	(3.498)	(0.254)	
Foreign nationality	0.021	0.057	0.036	0.133
	(0.142)	(0.232)	(0.016)	
Study program: Master level	0.231	0.214	-0.017	-0.029
	(0.422)	(0.411)	(0.031)	
Study program: Arts and humanities	0.268	0.197	-0.072	-0.120
	(0.443)	(0.398)	(0.030)	
Study program: Engineering	0.182	0.183	0.001	0.003
	(0.386)	(0.388)	(0.028)	
Study program: Natural sciences	0.093	0.127	0.034	0.076
	(0.291)	(0.333)	(0.026)	
Study program: Economics and business	0.272	0.288	0.016	0.025
	(0.445)	(0.454)	(0.034)	
N. of obs.	731	229	960	960

Notes: This table documents attrition at individual level in the second stage of the experiment. Attrition happens because, starting from all subjects entering the second stage, some cannot be matched due to a missing potential partner. In addition, we consider subjects as attrited if they are from a pair where one or both potential partners did not enter correctly their partner's random number. Column (1) shows means and standard deviation for non-attrited individuals. Column (2) shows means and standard deviation for attrited individuals. Column (3) shows estimated differences between attrited and non-attrited individuals and corresponding standard errors. Column (4) shows standardized differences.

Table A.4: Balancing Stage 2: Origin from Homogenous vs. Mixed Teams

	Males assigned to			Females assigned to		
	All-male teams	Mixed teams	p-value both equal	All-female teams	Mixed teams	p-value both equal
	(1)	(2)	(3)	(4)	(5)	(6)
A-level GPA	2.72	2.72	0.99	2.76	2.77	0.85
	(0.61)	(0.61)		(0.62)	(0.61)	
Top-tier high school	0.81	0.82	0.79	0.85	0.76	0.02
	(0.39)	(0.41)		(0.35)	(0.41)	
Age	22.67	22.53	0.68	22.65	22.71	0.84
	(3.28)	(3.00)		(2.84)	(3.00)	
Foreign nationality	0.03	0.02	0.55	0.02	0.02	0.98
	(0.16)	(0.13)		(0.13)	(0.13)	
Study program: Master level	0.27	0.24	0.61	0.19	0.22	0.52
	(0.44)	(0.42)		(0.39)	(0.42)	
Study program: Arts and humanities	0.21	0.24	0.53	0.34	0.29	0.26
	(0.41)	(0.44)		(0.48)	(0.44)	
Study program: Engineering	0.27	0.18	0.08	0.11	0.13	0.46
	(0.44)	(0.37)		(0.31)	(0.37)	
Study program: Natural sciences	0.10	0.11	0.69	0.09	0.08	0.64
	(0.29)	(0.29)		(0.29)	(0.29)	
Study program: Economics and business	0.30	0.32	0.63	0.25	0.22	0.51
	(0.46)	(0.44)		(0.43)	(0.44)	
N. of obs.	261	119	380	232	119	351

Notes: This table reports balancing checks for stage 2 regarding the subjects' origin from gender-homogenous and mixed first-stage teams. Columns (1) and (2) show means and standard deviation for males who were assigned to all-male or mixed teams, respectively. Column (3) shows p-values for tests of the hypothesis that the means are equal. Columns (4) to (6) report corresponding information for female subjects who were assigned to all-female or mixed teams, respectively.

Table A.5: Balancing Stage 2: Assignment to Potential Teammates

	Males assigned to			Females assigned to		
	Male potential teammate	Female potential teammate	p-value both equal	Female potential teammate	Male potential teammate	p-value both equal
	(1)	(2)	(3)	(4)	(5)	(6)
A-level GPA	2.75	2.68	0.28	2.75	2.78	0.59
	(0.62)	(0.62)		(0.59)	(0.62)	
Top-tier high school	0.83	0.81	0.63	0.80	0.84	0.43
	(0.38)	(0.38)		(0.40)	(0.38)	
Age	22.43	22.82	0.24	22.48	22.82	0.26
	(3.12)	(3.06)		(2.95)	(3.06)	
Foreign nationality	0.03	0.02	0.31	0.03	0.01	0.28
	(0.18)	(0.11)		(0.16)	(0.11)	
Master level	0.24	0.28	0.32	0.18	0.21	0.54
	(0.43)	(0.43)		(0.39)	(0.43)	
Arts and humanities	0.23	0.20	0.58	0.32	0.32	1.00
	(0.42)	(0.44)		(0.47)	(0.44)	
Engineering	0.25	0.24	0.77	0.11	0.12	0.66
	(0.43)	(0.38)		(0.31)	(0.38)	
Natural sciences	0.10	0.10	0.76	0.06	0.11	0.09
	(0.29)	(0.31)		(0.23)	(0.31)	
Econ. and business	0.29	0.31	0.62	0.24	0.24	0.89
	(0.46)	(0.45)		(0.43)	(0.45)	
N. of obs.	189	191	380	157	194	351

Notes: This table reports balancing checks for stage 2 regarding the subjects' assignment to female and male potential teammates. Columns (1) and (2) show means and standard deviation for males who were assigned to male or female potential teammates, respectively. Column (3) shows p-values for tests of the hypothesis that the means are equal. Columns (4) to (6) report corresponding information for female subjects.

Table A.6: Descriptives on Outcomes: Individual Level

	Mean (1)	Std. Dev. (2)
A. First-stage outcomes:		
Number of words	487.00	361.92
Number of turns	36.94	23.23
Own vocal semtiment: Positive	0.39	0.20
Own vocal semtiment: Negative	0.26	0.14
Perception: Positivity	4.64	0.64
Perception: Cooperativeness	4.65	0.64
Perception: Likeability	4.01	0.93
N. of obs.	1368	
B. Second-stage outcomes:		
Indicator: Subject prefers teamwork	0.80	0.40
Belief: Own productitivity	10.95	3.32
Belief: Partner's productivity	12.09	3.04
Belief: Team productitivity	14.73	2.95
Belief: Positivity	4.51	0.66
Belief: Cooperativeness	4.51	0.64
Belief: Likeability	4.09	0.85
N. of obs.	731	

Notes: This table shows descriptives for individual-level outcomes. In panel A, due to missing values in survey responses, the number of observations for the outcomes measuring perceptions varies between 1357 and 1362.

Table A.7: Descriptives on Outcomes: Team Level

	Mean (1)	Std. Dev. (2)
Number of problems solved	4.35	1.69
Number of words	1947.99	680.32
Number of turns	147.77	51.91
HHI words	0.34	0.06
HHI turns	0.31	0.04
Vocal semtiment: Positive	0.39	0.16
Vocal semtiment: Negative	0.25	0.11
Perception: Positivity	4.64	0.39
Perception: Cooperativeness	4.65	0.35
Perception: Likeability	4.01	0.57
N. of obs.	342	

Notes: This table shows descriptives for team-level outcomes.

Table A.8: Awareness of Team Gender Composition, First Stage

	= 1 if aware of exact team gender composition (1)	= 1 if aware of whether team is mixed or not (2)
Female (β_1)	-0.016	-0.015
	(0.019)	(0.019)
Mixed team (β_2)	-0.014	-0.014
	(0.020)	(0.020)
Female × Mixed team (β_3)	-0.106***	0.026
	(0.031)	(0.024)
N. of obs.	1352	1352
Mean dep. var.	0.94	0.96
Mean dep. var. all-male	0.97	0.97
Subject-level controls	Yes	Yes
$\beta_1 + \beta_3 = 0$ (p-value)	0.000	0.439
$\beta_2 + \beta_3 = 0$ (p-value)	0.000	0.532

Notes: This table shows OLS regressions using as dependent variables indicators for subjects who were aware of the team gender composition. In Column (1), we use an indicator for subjects whose answer to a survey question on how many of the teammates were female indicates awareness of the exact team gender composition. Column (2) adjusts the indicator by coding females in mixed teams as aware of the gender composition if their response suggests they counted themselves in when stating the number of female team members (the question asked for the number of females among the *other* team members). The regressions control for A-level GPA, age, A-level degree obtained from top-tier high school type, foreign nationality, study program at Master level, study field (arts and humanities, engineering, natural sciences, economics and business administration), and an indicator for teams where some members were silent during the team task. Standard errors (clustered at team level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Awareness of Potential Partner's Gender, Second Stage

	= 1 if subject is aware of potential partner's gender		
	All (1)	Females (2)	Males (3)
Female partner 2nd stage (β)	0.005	0.014	-0.001
	(0.012)	(0.017)	(0.014)
N. of obs.	731	351	380
Mean dependent variable	0.98	0.98	0.98
Subject-level controls	Yes	Yes	Yes

Notes: This table shows OLS regressions using as dependent variable an indicator for subjects who answered correctly a survey question on whether the potential partner in stage 2 was female. The regressions control for A-level GPA, age, and indicators for an A-level degree obtained from top-tier high school type, foreign nationality, study program at Master level, study field (arts and humanities, engineering, natural sciences, economics and business administration), and an indicator for teams where some members were silent during the team task. Column (1) also controls for gender. Standard errors (clustered at team level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Balancing Checks: Subjects Working Under Individual Piece Rate

	Males	Females	p-value both equal
	(1)	(2)	(3)
A-level GPA	2.70	2.75	0.47
	(0.62)	(0.57)	
Top-tier high school	0.81	0.80	0.84
	(0.39)	(0.40)	
Age	23.32	22.94	0.27
	(3.04)	(2.90)	
Study program: Master level	0.30	0.19	0.03
	(0.46)	(0.39)	
Foreign nationality	0.05	0.05	0.81
	(0.23)	(0.21)	
N. of obs.	149	147	296

Notes: This table reports balancing checks by gender for subjects who worked on the team task individually. Columns (1) and (2) show means and standard deviation for males and females, respectively. Column (3) shows p-values for tests of the hypothesis that the means are equal.

Table A.11: Gender Neutrality: Subjects Working Under Individual Piece Rate

	Number of problems solved	Likeability of the task
	(1)	(2)
Female	-0.121	-0.152
	(0.211)	(0.127)
A-level GPA	0.725***	-0.076
	(0.168)	(0.105)
Study program: Arts & humanities	0.103	0.220
	(0.288)	(0.176)
Study program: Engineering	0.300	0.300
	(0.334)	(0.187)
Study program: Natural sciences	-0.356	0.024
	(0.358)	(0.234)
Study program: Economics & business	-0.208	0.203
	(0.337)	(0.189)
Mean dep. var. males	4.46	3.21
N. of obs.	296	296
Subject-level controls	Yes	Yes

Notes: This table reports OLS regressions using the sample of subjects who worked on the team task under an individual piece rate (no communication with other subjects, no teamwork). Column (1) shows how the performance of subjects depends on gender, A-level GPA, and the series of study field indicators. In addition, the regressions control for A-level degree obtained from the top-tier high school type, age, study program at Master level, and foreign nationality. Column (2) reports an equivalent regression using as an outcome the subject's level of agreement with the statement "Working on the problems was fun" (5-point Likert scale, higher numbers indicating stronger agreement).

Table A.12: Beliefs About Potential Partner's Productivity

	Belief about partner's individual productivity		
	All	Females	Males
	(1)	(2)	(3)
Female partner 2nd stage (β)	0.212	0.084	0.333
	(0.262)	(0.384)	(0.344)
N. of obs.	731	351	380
Mean dependent variable	12.09	11.85	12.32
Subject-level controls	Yes	Yes	Yes
$\beta = 0$ (p-value MHT)		0.835	0.579

Notes: This table shows OLS regressions using as dependent variable the subjects' belief about the number of problems the potential partner would solve individually in a possible further task. All regressions control for gender (Column (1) only), A-level GPA, age, A-level degree obtained from top-tier high school type, foreign nationality, study program at Master level, study field (arts and humanities, engineering, natural sciences, economics and business administration), and an indicator for teams where some members were silent during the team task. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values adjusted for multiple hypothesis testing (MHT, two hypotheses included) follow Barsbai et al. (2020).

Table A.13: Robustness: Quantity of Communication Effects, Individual Level

	#Words	#Words	log(#Words)
	(1)	(2)	(3)
Female (β_1)	-73.21***	-81.18***	-0.20***
	(22.57)	(24.25)	(0.06)
Mixed team (β_2)	97.66***	99.10***	0.19***
	(29.09)	(28.07)	(0.07)
Female $\times$ Mixed team (β_3)	-169.41***	-182.98***	-0.41***
	(38.19)	(38.17)	(0.13)
A-level GPA		116.77***	0.30***
		(15.04)	(0.06)
Openness		2.89	
		(2.53)	
Conscientiousness		-1.69	
		(3.07)	
Extraversion		28.71***	
		(2.22)	
Agreeableness		-9.86***	
		(3.52)	
Neuroticism		7.73***	
		(2.52)	
Subject-level controls	No	Yes	Yes
Controls include Big 5	No	Yes	No
N. of obs.	1368	1281	1368
Adj. R^2	0.042	0.207	0.150
Mean dep. var. all-male	519.4	517.0	5.9
$\beta_1 + \beta_3 = 0$ (p-value)	0.000	0.000	0.000
$\beta_2 + \beta_3 = 0$ (p-value)	0.003	0.001	0.009

Notes: This table shows OLS regressions at the individual level using as dependent variables the number of words and the number of words in logs, respectively. Regressions control for age, A-level degree obtained from top-tier high school type, foreign nationality, study program at Master level, study field (arts and humanities, engineering, natural sciences, economics/business administration), and an indicator for teams with silent members. Standard errors (clustered at team level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.14: Effects on Total Speaking Time, Individual Level

	Total speaking time (in minutes)	
	(1)	(2)
Female (β_1)	-0.19	-0.21
	(0.14)	(0.15)
Mixed team (β_2)	0.71***	0.75***
	(0.17)	(0.16)
Female $\times$ Mixed team (β_3)	-1.12***	-1.18***
	(0.24)	(0.23)
A-level GPA	0.69***	0.70***
	(0.10)	(0.10)
Subject-level controls	Yes	Yes
Controls include Big 5	No	Yes
N. of obs.	1368	1281
Adj. R^2	0.086	0.186
Mean dep. var. all-male	3.25	3.24
$\beta_4 := \beta_1 + \beta_3$	-1.31	-1.40
$\beta_4 = 0$ (p-value)	0.000	0.000
$\beta_5 := \beta_2 + \beta_3$	-0.41	-0.44
$\beta_5 = 0$ (p-value)	0.013	0.008
$\beta_1 = 0$ (p-value MHT)	0.182	0.158
$\beta_2 = 0$ (p-value MHT)	0.000	0.000
$\beta_3 = 0$ (p-value MHT)	0.000	0.000

Notes: This table shows OLS regressions using as dependent variable the total speaking time at individual level. Regressions control for age, A-level degree obtained from top-tier high school type, foreign nationality, study program at Master level, study field (arts and humanities, engineering, natural sciences, economics and business administration), and an indicator for subjects from teams with silent members. Column (2) additionally controls for the Big 5 personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism). Standard errors (clustered at team level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values adjusted for multiple hypothesis testing (MHT) follow Barsbai et al. (2020). Multiple testing is done separately by column (three hypotheses in each regression).

Table A.15: List of Topic Words

	Problem set A		Problem set B	
	Word	%	Word	%
1	D	14.49	B	18.99
2	C	13.31	C	10.75
3	B	11.44	D	9.98
4	A	8.23	A	8.52
5	market	5.26	invest	4.62
6	drug	3.35	investment	4.15
7	doctor	2.70	rise	3.78
8	market share	2.64	innovation capital	3.54
9	sales	2.44	country	2.16
10	company	2.35	development	2.01
11	emerging	2.31	human capital	1.97
12	rise	2.27	APPNAME	1.84
13	country	2.03	capital	1.71
14	prescription	1.83	knowledge capital	1.51
15	growth	1.62	physical	1.43
16	performance	1.58	company	1.32
17	market access	1.53	innovation	1.22
18	North America	1.52	investor	1.02
19	year	1.35	conviction	1.02
20	COMPANYNAME	1.34	price	0.98
21	health insurance	1.00	networking	0.96
22	tobacco	0.94	economic	0.90
23	profit margin	0.93	app	0.84
24	profit	0.89	awareness	0.81
25	growth opportunity	0.80	event	0.81
26	patent protection	0.79	type	0.79
27	bribe	0.67	social	0.79
28	invest	0.63	brand value	0.78
29	pay	0.59	productivity growth	0.74
30	vaccination campaign	0.58	market share	0.71
31	medicine	0.58	product	0.71
32	alcohol consumption	0.54	industry	0.69
33	future	0.51	profit margin	0.69
34	competitor	0.48	database	0.68
35	alcohol	0.48	productivity	0.58
36	management	0.47	difference	0.51
37	change	0.47	asset value	0.46
38	pharmaceuticals	0.46	design concept	0.45
39	traditional	0.46	organization	0.42
40	herbal	0.42	COMPANYNAME	0.42
41	self-medication	0.42	training programs	0.40
42	disease	0.40	military	0.40
43	obstacle	0.39	activity	0.38
44	female doctor	0.39	large enterprise	0.38
45	medicine	0.39	software	0.38
46	trend	0.38	management	0.37
47	income	0.37	authoritarian	0.37
48	investment	0.35	technology	0.36
49	national language	0.33	emigration wave	0.35
50	government output	0.33	collaboration	0.34
	Total	100.00	Total	100.00

Notes: This table shows all words from the team conversations (translated from German) we considered when defining the set of topic words. The inclusion of "A", "B", "C", and "D" accounts for references to the four possible solutions to each problem, which were labeled from *a* to *d*. For each problem set, we pre-selected from the information materials and problems all words that are topically related to the task and would unlikely be used in a conversation unrelated to it. The columns showing shares report how often a given word was used in relation to all listed words. The analyses reported in the paper are based on the 10 most frequently used topic words in each problem set. We use lists of topic words comprising the 20, 30, 40, or 50 most frequently used words in several robustness checks.

Table A.16: Robustness: Effects on #Topic Words, Team Level

	#Topic words				
	Number of topic words considered				
	10	20	30	40	50
	(1)	(2)	(3)	(4)	(5)
Gender-mixed team (β_1)	-12.2**	-19.3***	-21.8***	-24.2***	-26.0***
	(4.7)	(6.6)	(7.4)	(8.1)	(8.7)
All-female team (β_2)	-20.2***	-29.5***	-32.3***	-35.1***	-37.1***
	(5.2)	(7.2)	(8.0)	(8.8)	(9.4)
N. of obs.	342	342	342	342	342
Team-level controls	Yes	Yes	Yes	Yes	Yes
Mean dep. var. all-male	127.3	159.3	174.7	185.3	192.8
$\beta_1 = \beta_2$ (p-value)	0.093	0.119	0.150	0.175	0.193

Notes: This table shows OLS regressions at team level. The regressions differ by the definition of the dependent variable, capturing the number of topic words (i.e., words that are topically related to the team task). Column (1) defines as topic words only the 10 most frequent words that are topically related to the task and thus repeats the regression shown in Table 4, Column (3). The remaining columns consider more broadly defined sets of topic words. Column (5) uses all words on the list provided in Appendix Table A.15. All regressions control for team averages of A-level GPA and age, maximum and minimum A-level GPA, maximum and minimum age, the share of team members with an A-level degree obtained from top-tier high school type, the share of team members with foreign nationality, the share of team members studying at Master level, a series of variables capturing the shares of team members studying in one of the main study fields (arts and humanities, engineering, natural sciences, economics/business administration), and an indicator for teams where some members were silent during the team task. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.17: Splitting up the Mixed-Team Effect by Teams' GPA-by-Gender Composition

	#Problems solved (1)	#Words (2)
Mixed team: females below, males above median (β_1)	-0.70**	-235.87*
	(0.30)	(123.37)
Mixed team: females above, males below median (β_2)	0.07	-92.19
	(0.30)	(116.44)
Mixed team: one female and one male above median (β_3)	-0.62*	-78.86
	(0.33)	(124.98)
All-female team (β_4)	-0.58**	-297.07***
	(0.25)	(95.25)
N. of obs.	342	342
Mean dep. var. all-male	4.6	2077.7
Team-level controls	Yes	Yes
$\beta_1 = \beta_2 = \beta_3 = \beta_4$ (p-value)	0.130	0.202

Notes: This table shows OLS regressions at the team level. The dependent variables are the number of problems solved (Column 1) and the word count variable (Column 2). The regressions use three different indicator variables for mixed teams, capturing the different possible team compositions by GPA: mixed teams with both females below median GPA and both males above median GPA (β_1), mixed teams with both females above median GPA and both males below median GPA (β_2), and mixed teams with one female and one male below median GPA and one female and one male above median GPA (β_3). All regressions control for team averages of A-level GPA and age, maximum and minimum A-level GPA, maximum and minimum age, the share of team members with an A-level degree obtained from top-tier high school type, the share of team members with foreign nationality, the share of team members studying at Master level, a series of variables capturing the shares of team members studying in one of the main study fields (arts and humanities, engineering, natural sciences, economics/business administration), and an indicator for teams where some members were silent during the team task. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.18: Performance: Teams vs. Individuals

	Number of problems solved (1)	(2)
Teamwork	-0.003	0.444***
	(0.136)	(0.154)
Constant	4.351***	4.351***
	(0.101)	(0.101)
N. of obs.	638	496
Teams with imperfect coordination excluded	No	Yes

Notes: This table shows an OLS regression that jointly uses team-level observations and observations from individuals working under an individual piece rate and regresses the number of correctly solved problems on an indicator for teams. Column (1) includes all observations (342 teams and 296 individuals). Column (2) includes all individuals, and in addition all teams that successfully coordinated their answers in all 10 problems (i.e., teams where all team members gave identical answers to all problems). No controls included. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.19: Coordination Within Teams

	#problems with perfect coordination (1)	Teams with perfect coordination: #problems solved (2)
Gender-mixed team (β_1)	-0.096	-0.436
	(0.180)	(0.274)
All-female team (β_2)	0.099	-0.565*
	(0.151)	(0.325)
N. of obs.	342	200
Team-level controls	Yes	Yes
Mean dep. var. all-male	9.29	5.04
$\beta_1 = \beta_2$ (p-value)	0.237	0.680

Notes: This table shows in Column (1) an OLS regression using as dependent variable the number of problems with perfect coordination among team members. Column (2) uses as dependent variable our measure of team performance (number of problems solved), but uses only teams that perfectly coordinated their answers in all 10 problems. All regressions control for team averages of A-level GPA and age, maximum and minimum A-level GPA, maximum and minimum age, the share of team members with an A-level degree obtained from top-tier high school type, the share of team members with foreign nationality, the share of team members studying at Master level, a series of variables capturing the shares of team members studying in one of the main study fields (arts and humanities, engineering, natural sciences, economics/business administration), and an indicator for teams where some members were silent during the team task. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.20: Robustness: Quantity of Communication and Team Performance

	Number of problems solved				
	Number of topic words considered				
	10	20	30	40	50
	(1)	(2)	(3)	(4)	(5)
#all words (β_1)	-0.001**	-0.001**	-0.001**	-0.001**	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
#topic words (β_2)	0.015***	0.011***	0.010***	0.010***	0.008***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
N. of obs.	342	342	342	342	342
Mean dep. var.	4.35	4.35	4.35	4.35	4.35
Team-level controls	Yes	Yes	Yes	Yes	Yes

Notes: This table shows OLS regressions using as dependent variable the number of problems solved at team level. The regressions do not condition on team gender composition but use as regressors of interest the overall number of words and the number of words that are topically related to the team task. The regressions differ by the definition of the latter variable. Column (1) defines as topic words only the 10 most frequent words that are topically related to the task, and thus repeats the regression shown in Table 6. The remaining columns consider more broadly defined sets of topic words. Column (5) uses all words on the list provided in Appendix Table A.15. All regressions control for team averages of A-level GPA and age, maximum and minimum A-level GPA and age, the share of team members with an A-level degree from the top-tier high school type, the share of team members with foreign nationality, the share of team members studying at Master level, a series of variables capturing the shares of team members studying in one of the main study fields (arts and humanities, engineering, natural sciences, economics/business administration), and an indicator for teams where some members were silent during the team task. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.21: Robustness: No Gender Gap in Share of Topic Words

	Share of topic words				
	Number of topic words considered				
	10	20	30	40	50
	(1)	(2)	(3)	(4)	(5)
Female (β_1)	0.001	-0.002	-0.002	-0.002	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mixed team (β_2)	0.000	-0.000	-0.000	-0.000	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Female $\times$ Mixed team (β_3)	0.001	0.000	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
A-level GPA	-0.001	0.001	0.002	0.002*	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N. of obs.	1336	1336	1336	1336	1336
Mean dep. var. all-male	0.065	0.079	0.085	0.089	0.093
Subject-level controls	Yes	Yes	Yes	Yes	Yes
$\beta_1 + \beta_3 = 0$ (p-value)	0.538	0.541	0.319	0.188	0.148
$\beta_2 + \beta_3 = 0$ (p-value)	0.708	0.996	0.777	0.606	0.555

Notes: This table shows subject-level OLS regressions using as dependent variable the share of words in a subject's utterances that are topically related to the team task. The regressions differ by the definition of topic words. Column (1) defines as topic words only the 10 most frequent words that are topically related to the task, and thus repeats the regression shown in Table 7. The remaining columns consider more broadly defined sets of topic words. Column (5) uses all words on the list provided in Appendix Table A.15. All Regressions control for age, A-level degree obtained from top-tier high school type, foreign nationality, study program at Master level, study field (arts and humanities, engineering, natural sciences, economics/business administration) and an indicator for teams with silent members. Standard errors (clustered at team level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.22: Distributional Effects on Team Communication

	HHI words	HHI turns
	(1)	(2)
Gender-mixed team (β_1)	0.013	0.007
	(0.009)	(0.005)
All-female team (β_2)	-0.007	-0.002
	(0.008)	(0.005)
N. of obs.	342	342
Mean dep. var. all-male	0.34	0.31
Team-level controls	Yes	Yes
$\beta_1 = \beta_2$ (p-value)	0.017	0.072
$\beta_1 = 0$ (p-value MHT)	0.365	0.351
$\beta_2 = 0$ (p-value MHT)	0.547	0.666

Notes: This table shows OLS regressions using as dependent variables the HHI of the number of words and the HHI of the number of turns at the team level, respectively. All regressions control for team averages of A-level GPA and age, maximum and minimum A-level GPA, maximum and minimum age, the share of team members with an A-level degree obtained from the top-tier high school type, the share of team members with foreign nationality, the share of team members studying at Master level, a series of variables capturing the shares of team members studying in one of the main study fields (arts and humanities, engineering, natural sciences, economics/business administration), and an indicator for teams where some members were silent during the team task. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values adjusted for multiple hypothesis testing (MHT, four hypotheses included) follow Barsbai et al. (2020).

Table A.23: Effects on Sentiment, Team Level

	Positive	Negative
	(1)	(2)
Gender-mixed team (β_1)	0.088***	-0.008
	(0.017)	(0.015)
All-female team (β_2)	0.254***	-0.063***
	(0.017)	(0.015)
N. of obs.	342	342
Mean dep. var. all-male	0.27	0.27
Team-level controls	Yes	Yes
$\beta_1 = \beta_2$ (p-value)	0.000	0.000
$\beta_1 = 0$ (p-value MHT)	0.000	0.605
$\beta_2 = 0$ (p-value MHT)	0.000	0.000

Notes: This table shows OLS regressions using as dependent variables measures of the sentiment of team communication captured through vocal features. Positive (negative) sentiment captures vocal features indicating happiness (sadness). All regressions control for team averages of A-level GPA and age, maximum and minimum A-level GPA, maximum and minimum age, the share of team members with an A-level degree obtained from top-tier high school type, the share of team members with foreign nationality, the share of team members studying at Master level, a series of variables capturing the shares of team members studying in one of the main study fields (arts and humanities, engineering, natural sciences, economics/business administration), and an indicator for teams where some members were silent during the team task. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values adjusted for multiple hypothesis testing (MHT, four hypotheses included) follow Barsbai et al. (2020).

Table A.24: Effects on Perceived Team Interaction

	Positivity	Cooperativeness	Likeability
	(1)	(2)	(3)
Gender-mixed team (β_1)	-0.029	-0.017	-0.021
	(0.051)	(0.046)	(0.077)
All-female team (β_2)	-0.034	-0.004	-0.113
	(0.057)	(0.051)	(0.081)
N. of obs.	342	342	342
Mean dep. var. all-male	4.65	4.66	4.06
Team-level controls	Yes	Yes	Yes
$\beta_1 = \beta_2$ (p-value)	0.929	0.797	0.253
$\beta_1 = 0$ (p-value MHT)	0.952	0.976	0.958
$\beta_2 = 0$ (p-value MHT)	0.971	0.948	0.556

Notes: This table shows OLS regressions using as dependent variables measures of perceived team communication. Perceived positivity, cooperativeness, and likeability of the team task are all measured using a 5-point Likert scale. All regressions control for team averages of A-level GPA and age, maximum and minimum A-level GPA, maximum and minimum age, the share of team members with an A-level degree obtained from top-tier high school type, the share of team members with foreign nationality, the share of team members studying at Master level, a series of variables capturing the shares of team members studying in one of the main study fields (arts and humanities, engineering, natural sciences, economics/business administration), and an indicator for teams where some members were silent during the team task. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values adjusted for multiple hypothesis testing (MHT, six hypotheses included) follow Barsbai et al. (2020).

Table A.25: Perceived Communication: Secondary Outcomes, Individual Level

	Sufficient communication (1)	Symmetric communication (2)	Letting others finish (3)
Female (β_1)	-0.049	0.171**	-0.029
	(0.067)	(0.085)	(0.042)
Mixed team (β_2)	-0.091	-0.119	-0.036
	(0.080)	(0.100)	(0.047)
Female $\times$ Mixed team (β_3)	0.022	-0.084	0.025
	(0.104)	(0.124)	(0.070)
N. of obs.	1357	1362	1357
Mean dep. var. all-male	4.29	3.31	4.71
Subject-level controls	Yes	Yes	Yes
$\beta_4 := \beta_1 + \beta_3$	-0.027	0.087	-0.003
$\beta_4 = 0$ (p-value)	0.737	0.344	0.950
$\beta_5 := \beta_2 + \beta_3$	-0.069	-0.203	-0.011
$\beta_5 = 0$ (p-value)	0.457	0.045	0.846
$\beta_1 = 0$ (p-value MHT)	0.935	0.296	0.853
$\beta_2 = 0$ (p-value MHT)	0.825	0.814	0.941
$\beta_3 = 0$ (p-value MHT)	0.828	0.916	0.921

Notes: This table shows OLS regressions using as dependent variables measures of individual perceptions of team communication. Perceptions of whether the team communicated sufficiently and symmetric and whether the team members let each other finish are all measured using a 5-point Likert scale. All regressions control for A-level GPA, age, and indicators for an A-level degree obtained from top-tier high school type, foreign nationality, study program at Master level, study field (arts and humanities, engineering, natural sciences, economics and business administration), and an indicator for teams where some members were silent during the team task. Standard errors (clustered at team level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.26: Perceived Communication: Secondary Outcomes, Team Level

	Sufficient communication (1)	Symmetric communication (2)	Letting others finish (3)
Gender-mixed team (β_1)	-0.126	-0.065	-0.038
	(0.077)	(0.094)	(0.043)
All-female team (β_2)	-0.078	0.182**	-0.050
	(0.075)	(0.092)	(0.045)
N. of obs.	342	342	342
Mean dep. var. all-male	4.29	3.31	4.71
Team-level controls	Yes	Yes	Yes
$\beta_1 = \beta_2$ (p-value)	0.552	0.007	0.806
$\beta_1 = 0$ (p-value MHT)	0.428	0.511	0.607
$\beta_2 = 0$ (p-value MHT)	0.672	0.239	0.713

Notes: This table shows OLS regressions using as dependent variables measures of perceived team communication. All outcomes are measured using a 5-point Likert scale. All regressions control for team averages of A-level GPA and age, maximum and minimum A-level GPA, maximum and minimum age, the share of team members with an A-level degree obtained from top-tier high school type, the share of team members with foreign nationality, the share of team members studying at Master level, a series of variables capturing the shares of team members studying in one of the main study fields (arts and humanities, engineering, natural sciences, economics/business administration), and an indicator for teams where some members were silent during the team task. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values adjusted for multiple hypothesis testing (MHT, six hypotheses included) follow Barsbai et al. (2020).

Table A.27: Productivity Beliefs: Past Exposure to Mixed Teamwork

	Belief about productivity:					
	Own		Partner		Team	
	(1)	(2)	(3)	(4)	(5)	(6)
Female (β_1)	-1.290***	-1.348***	-0.373	-0.458	-0.650***	-0.600**
	(0.270)	(0.333)	(0.260)	(0.326)	(0.247)	(0.301)
Mixed team (β_2)	0.322	0.239	0.321	0.199	0.250	0.322
	(0.262)	(0.400)	(0.264)	(0.381)	(0.239)	(0.379)
Female $\times$ Mixed team (β_3)		0.171		0.252		-0.150
		(0.564)		(0.518)		(0.528)
N. of obs.	731	731	731	731	731	731
Mean dep. var. all-male	11.55	11.55	12.26	12.26	15.00	15.00
Subject-level controls	Yes	Yes	Yes	Yes	Yes	Yes
$\beta_4 := \beta_1 + \beta_3$		-1.176		-0.206		-0.750
$\beta_4 = 0$ (p-value)		0.011		0.617		0.086
$\beta_5 := \beta_2 + \beta_3$		0.410		0.451		0.172
$\beta_5 = 0$ (p-value)		0.269		0.211		0.604
$\beta_1 = 0$ (p-value MHT)	0.000	0.000	0.388	0.495	0.045	0.214
$\beta_2 = 0$ (p-value MHT)	0.382	0.917	0.341	0.933	0.309	0.807
$\beta_3 = 0$ (p-value MHT)		0.914		0.881		0.779

Notes: This table shows OLS regressions using as dependent variables different measures of beliefs about productivity in a possible further task. Columns (1) and (2) analyze beliefs about a subject's own productivity if working on the task individually. Columns (3) and (4) study subjects' beliefs about the potential partner's individual productivity. Columns (5) and (6) consider beliefs about team productivity in case of joint work with the potential partner. All regressions control for A-level GPA, age, A-level degree obtained from top-tier high school type, foreign nationality, study program at Master level, study field (arts and humanities, engineering, natural sciences, economics and business administration), and an indicator for teams where some members were silent during the team task. Standard errors (in parentheses) account for clusters comprising all subjects from first-stage teams used in the cross-wise random assignment to pairs of potential partners. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values adjusted for multiple hypothesis testing (MHT) follow Barsbai et al. (2020). Multiple testing is done across Columns (1), (3), and (5) (six hypotheses) and across Columns (2), (4), and (6) (nine hypotheses), respectively.

Table A.28: Effects on Uncertainty in Speech

	Incidence of uncertainty phrases
Female (β_1)	0.219***
	(0.046)
Mixed team (β_2)	0.045
	(0.039)
Female × Mixed team (β_3)	-0.031
	(0.078)
A-level GPA	-0.067*
	(0.034)
N. of obs.	1336
Mean dep. var. all-male	0.477
Subject-level controls	Yes
$\beta_4 := \beta_1 + \beta_3$	0.188
$\beta_4 = 0$ (p-value)	0.004
$\beta_5 := \beta_2 + \beta_3$	0.014
$\beta_5 = 0$ (p-value)	0.844
$\beta_1 = 0$ (p-value MHT)	0.000
$\beta_2 = 0$ (p-value MHT)	0.412
$\beta_3 = 0$ (p-value MHT)	0.702

Notes: This table shows an OLS regression using as dependent variable the incidence of uncertainty phrases (number of such phrases per 100 words) at individual level. Uncertainty phrases are defined by the occurrence of the following combination of words in a sentence: "I + not + sure", "I + uncertain", "I + waver", "I + not + know", "I + not + understand", "could + be", "no + idea", "unsettle", and "unclear". Regressions control for age, A-level degree obtained from top-tier high school type, foreign nationality, study program at Master level, study field (arts and humanities, engineering, natural sciences, economics and business administration), and an indicator for subjects from teams with silent members. Standard errors (clustered at team level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values adjusted for multiple hypothesis testing (MHT, three hypotheses included) follow Barsbai et al. (2020).

Table A.29: Productivity Beliefs: Past Exposure and Partner's Gender

	Belief about productivity:					
	Own		Partner		Team	
	Females (1)	Males (2)	Females (3)	Males (4)	Females (5)	Males (6)
Female partner 2nd stage (β_1)	0.303	0.285	-0.238	0.357	0.002	0.467
	(0.494)	(0.439)	(0.482)	(0.417)	(0.441)	(0.413)
Mixed team 1st stage (β_2)	0.504	0.149	0.141	0.259	0.318	0.383
	(0.546)	(0.553)	(0.561)	(0.478)	(0.472)	(0.523)
Female partner 2nd stage $\times$ Mixed team 1st stage (β_3)	-0.317	0.326	0.713	0.020	-0.194	0.016
	(0.712)	(0.790)	(0.736)	(0.751)	(0.645)	(0.720)
N. of obs.	351	380	351	380	351	380
Mean dep. var. gender-homogenous teams	10.07	11.55	11.69	12.26	14.27	15.00
Subject-level controls	Yes	Yes	Yes	Yes	Yes	Yes
$\beta_4 := \beta_1 + \beta_3$	-0.014	0.611	0.474	0.377	-0.192	0.482
$\beta_4 = 0$ (p-value)	0.979	0.352	0.414	0.549	0.708	0.415
$\beta_5 := \beta_2 + \beta_3$	0.187	0.475	0.854	0.278	0.124	0.399
$\beta_5 = 0$ (p-value)	0.693	0.424	0.077	0.644	0.785	0.461
$\beta_1 = 0$ (p-value MHT)	0.999	0.998	1.000	0.994	0.996	0.956
$\beta_2 = 0$ (p-value MHT)	0.980	1.000	0.997	1.000	0.998	0.996
$\beta_3 = 0$ (p-value MHT)	1.000	0.999	0.974	1.000	1.000	1.000

Notes: This table shows OLS regressions using as dependent variables different measures of beliefs about productivity in a possible further task. Columns (1) and (2) analyze beliefs about a subject's own productivity if working on the task individually. Columns (3) and (4) study subjects' beliefs about the potential partner's individual productivity. Columns (5) and (6) consider beliefs about team productivity in case of joint work with the potential partner. All regressions control for A-level GPA, age, A-level degree obtained from top-tier high school type, foreign nationality, study program at Master level, study field (arts and humanities, engineering, natural sciences, economics and business administration), and an indicator for teams where some members were silent during the team task. Standard errors (in parentheses) account for clusters comprising all subjects from first-stage teams used in the cross-wise random assignment to pairs of potential partners. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values adjusted for multiple hypothesis testing (MHT, 18 hypotheses included) follow Barsbai et al. (2020).

Table A.30: Communication-Related Beliefs: Past Exposure

	Belief about:							
	Positivity		Cooperativeness		Likeability		Belief index	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female (β_1)	0.059	0.051	-0.060	-0.000	0.057	0.059	0.022	0.054
	(0.051)	(0.067)	(0.050)	(0.060)	(0.067)	(0.089)	(0.078)	(0.102)
Mixed team (β_2)	0.132***	0.121*	0.090*	0.177**	0.052	0.054	0.151*	0.199*
	(0.044)	(0.064)	(0.053)	(0.070)	(0.070)	(0.105)	(0.077)	(0.108)
Female × Mixed team (β_3)		0.022		-0.179*		-0.005		-0.098
		(0.108)		(0.100)		(0.155)		(0.168)
N. of obs.	731	731	731	731	731	731	731	731
Mean dep. var. all-male	4.45	4.45	4.49	4.49	4.07	4.07	-0.00	-0.00
Subject-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\beta_4 \coloneqq \beta_1 + \beta_3$		0.073		-0.179		0.054		-0.044
$\beta_4 = 0$ (p-value)		0.371		0.033		0.642		0.732
$\beta_5 \coloneqq \beta_2 + \beta_3$		0.144		-0.002		0.049		0.100
$\beta_5 = 0$ (p-value)		0.057		0.975		0.634		0.402
$\beta_1 = 0$ (p-value MHT)	0.523	0.909	0.582	0.996	0.635	0.927	0.785	0.587
$\beta_2 = 0$ (p-value MHT)	0.016	0.285	0.320	0.073	0.471	0.960	0.110	0.149
$\beta_3 = 0$ (p-value MHT)		0.995		0.294		1.000		0.759

Notes: This table shows OLS regressions using as dependent variables beliefs about team communication team in a possible further team interaction with the potential partner. Beliefs about positivity, cooperativeness, and likeability of the team task are all measured using a 5-point Likert scale. The belief index is constructed by aggregating standardized beliefs in all three dimensions (Kling et al., 2007). All regressions control for A-level GPA, age, A-level degree obtained from top-tier high school type, foreign nationality, study program at Master level, study field (arts and humanities, engineering, natural sciences, economics and business administration), and an indicator for teams where some members were silent during the team task. Standard errors (in parentheses) account for clusters comprising all subjects from first-stage teams used in the cross-wise random assignment to pairs of potential partners. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values adjusted for multiple hypothesis testing (MHT) follow Barsbai et al. (2020). Multiple testing is done across Columns (1), (3), and (5) (six hypotheses), across Columns (2), (4), and (6) (nine hypotheses), and separately for Columns (7) (two hypotheses) and (8) (three hypotheses).

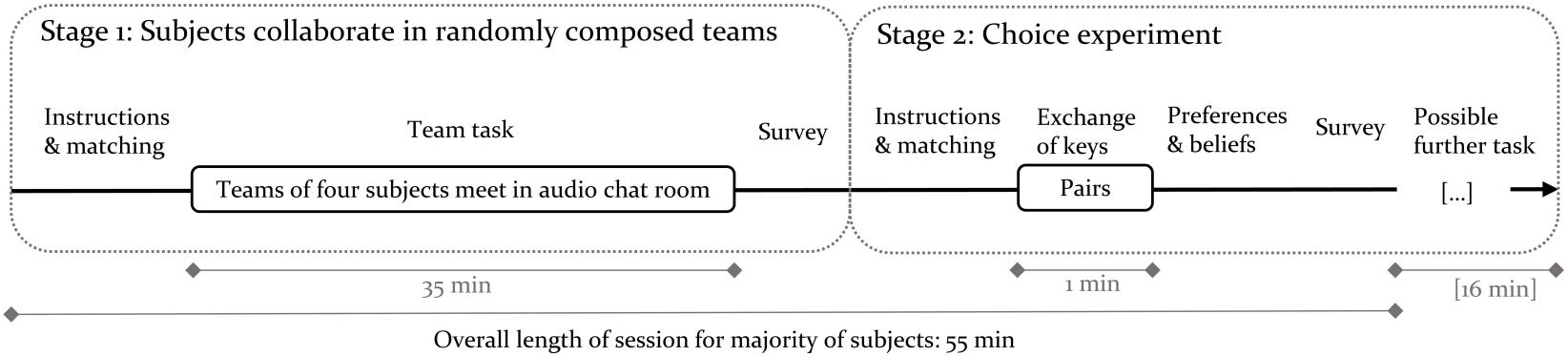
Table A.31: Communication-Related Beliefs: Past Exposure and Partner's Gender

	Belief about:							
	Positivity		Cooperativeness		Likeability		Belief index	
	Females (1)	Males (2)	Females (3)	Males (4)	Females (5)	Males (6)	Females (7)	Males (8)
Female partner 2nd stage (β_1)	0.315***	0.153	0.355***	0.033	0.346***	0.154	0.544***	0.173
	(0.081)	(0.093)	(0.074)	(0.089)	(0.126)	(0.124)	(0.128)	(0.144)
Mixed team 1st stage (β_2)	0.218*	0.187*	0.179	0.188*	0.006	0.043	0.232	0.237
	(0.113)	(0.099)	(0.116)	(0.104)	(0.136)	(0.143)	(0.176)	(0.158)
Female partner 2nd stage × Mixed team 1st stage (β_3)	-0.227	-0.104	-0.435***	0.002	-0.034	0.093	-0.404*	-0.015
	(0.141)	(0.146)	(0.154)	(0.153)	(0.192)	(0.200)	(0.229)	(0.227)
N. of obs.	351	380	351	380	351	380	351	380
Mean dep. var. gender-homogenous teams	4.48	4.45	4.48	4.49	4.09	4.07	0.02	-0.00
Subject-level controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\beta_4 \coloneqq \beta_1 + \beta_3$	0.089	0.049	-0.079	0.034	0.312	0.248	0.140	0.158
$\beta_4 = 0$ (p-value)	0.437	0.671	0.556	0.772	0.025	0.126	0.452	0.374
$\beta_5 \coloneqq \beta_2 + \beta_3$	-0.009	0.083	-0.256	0.190	-0.028	0.137	-0.172	0.222
$\beta_5 = 0$ (p-value)	0.919	0.385	0.009	0.068	0.853	0.342	0.267	0.150
$\beta_1 = 0$ (p-value MHT)	0.011	0.567	0.000	0.995	0.139	0.774	0.000	0.369
$\beta_2 = 0$ (p-value MHT)	0.447	0.453	0.588	0.481	0.999	0.996	0.445	0.398
$\beta_3 = 0$ (p-value MHT)	0.563	0.973	0.086	0.988	0.998	0.996	0.296	0.953

Notes: This table shows OLS regressions using as dependent variables beliefs about team communication team in a possible further team interaction with the potential partner. Beliefs about positivity, cooperativeness, and likeability of the team task are all measured using a 5-point Likert scale. The belief index is constructed by aggregating standardized beliefs in all three dimensions (Kling et al., 2007). All regressions control for A-level GPA, age, A-level degree obtained from top-tier high school type, foreign nationality, study program at Master level, study field (arts and humanities, engineering, natural sciences, economics and business administration), and an indicator for teams where some members were silent during the team task. Standard errors (in parentheses) account for clusters comprising all subjects from first-stage teams used in the cross-wise random assignment to pairs of potential partners. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values adjusted for multiple hypothesis testing (MHT) follow Barsbai et al. (2020). Multiple testing is done across Columns (1) to (6) (18 hypotheses), and across Columns (7) and (8) (six hypotheses).

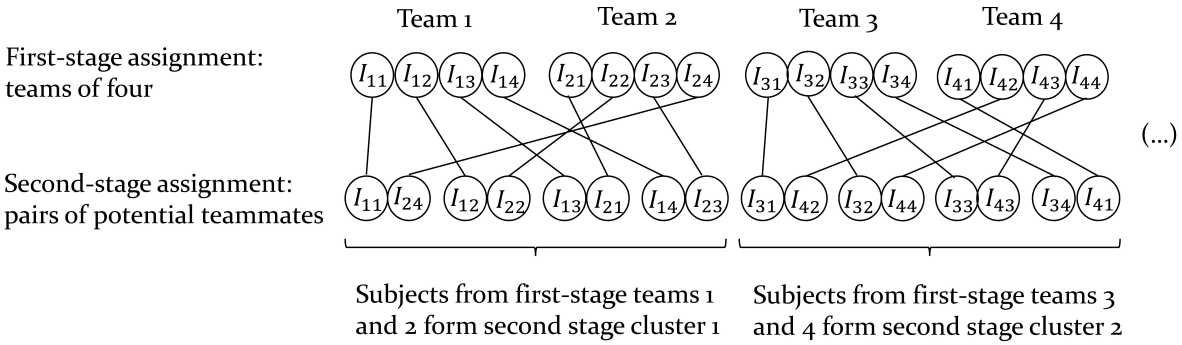
B Appendix Figures

Figure B.1: Timeline of Experimental Design



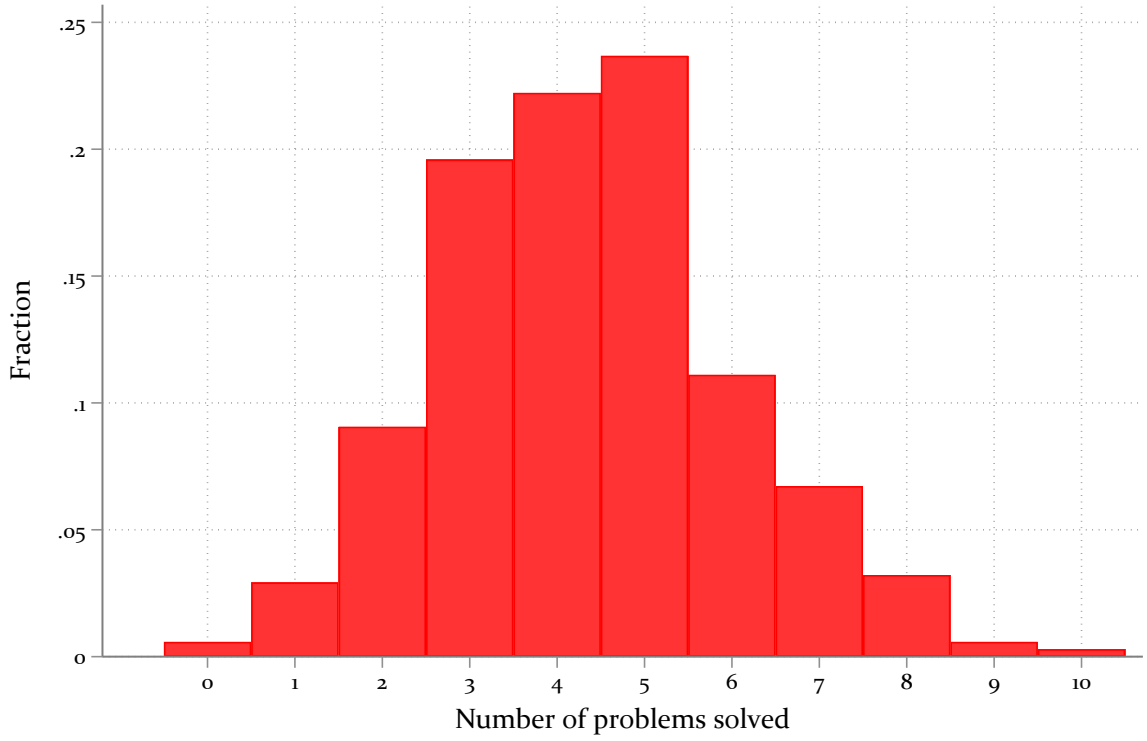
Notes: This figure shows the timeline of the experimental design. After initial instructions, in stage 1 randomly composed teams of four subjects met in an audio chat room and worked on a team task for 30 minutes (35 minutes including extra time for reading information material). Subjects then answered a survey individually. After further instructions, in stage 2 each subject met a randomly selected other subject in the audio chat room for one minute to exchange a private key. Subjects then answered a survey on preferences and beliefs individually. For the majority of subjects, the experiment took about 55 minutes. For some randomly selected second-stage pairs, a further real-effort task followed. Completing this task took another 16 minutes.

Figure B.2: Graphical Illustration of Second-Stage Matching



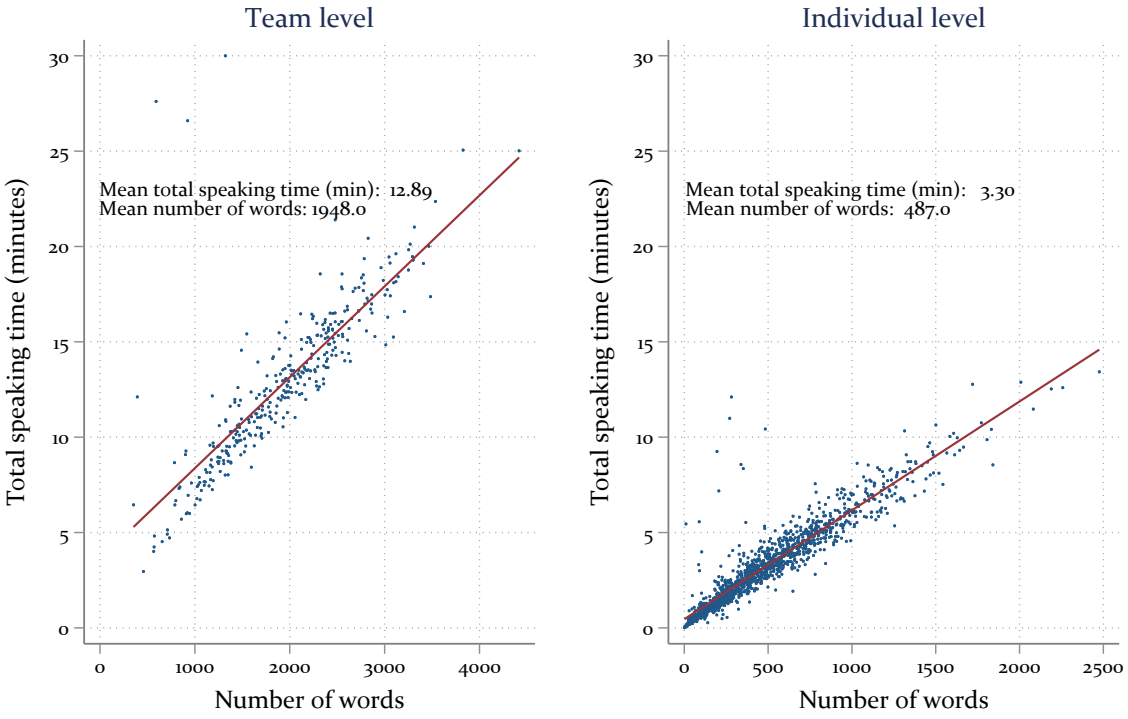
Notes: This figure illustrates the matching of subjects in stage 2 of the experimental design. The matching was based on a random formation of first-stage team pairs. In each pair of first-stage teams, subjects were randomly matched with a subject from the other team. As a result, all subjects were matched with a randomly selected stranger. Second-stage clusters comprise all subjects from the respective first-stage team pairs. In the case of an odd number of first-stage teams, one second-stage cluster comprised the subjects from three first-stage teams.

Figure B.3: Histogram of Number of Problems Solved



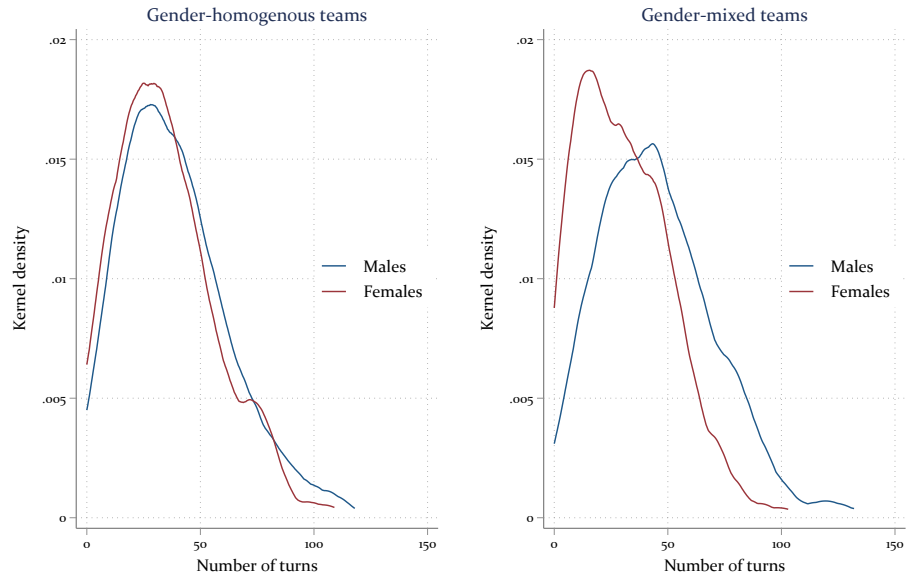
Notes: This figure shows a histogram of number of problems solved. The sample consists of all teams ($N = 342$).

Figure B.4: Number of Words vs. Total Speaking Time



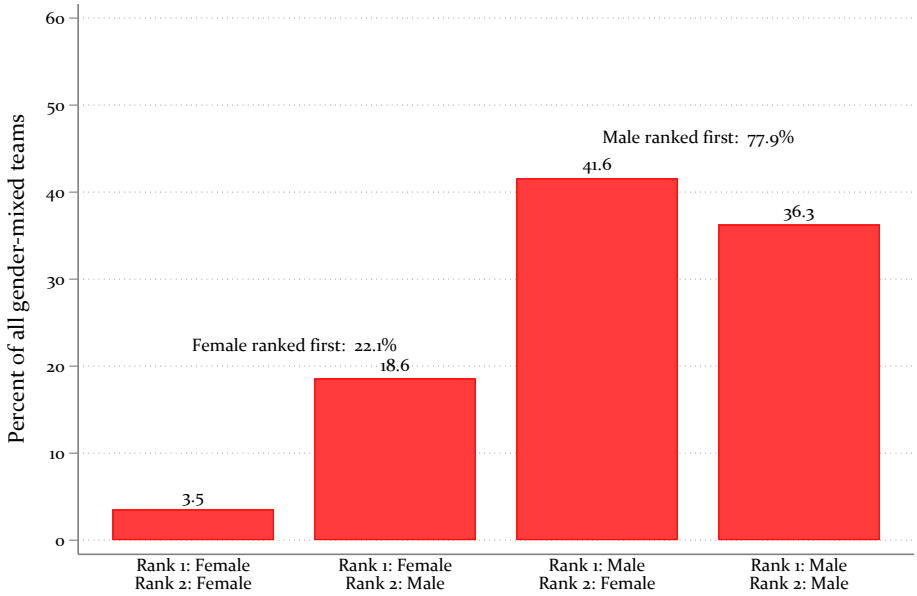
Notes: This figure shows plots of total speaking time against number of words, separately for team and individual level. Since we measure speaking time based on an algorithm that removes periods of silence from the audio recordings, speaking time tends to be overstated in case of background noise, leading to outliers. The team-level plot is based on all 342 teams. The individual-level plot uses the data from all 1386 subjects in these teams.

Figure B.5: Total Number of Turns, Individual Level



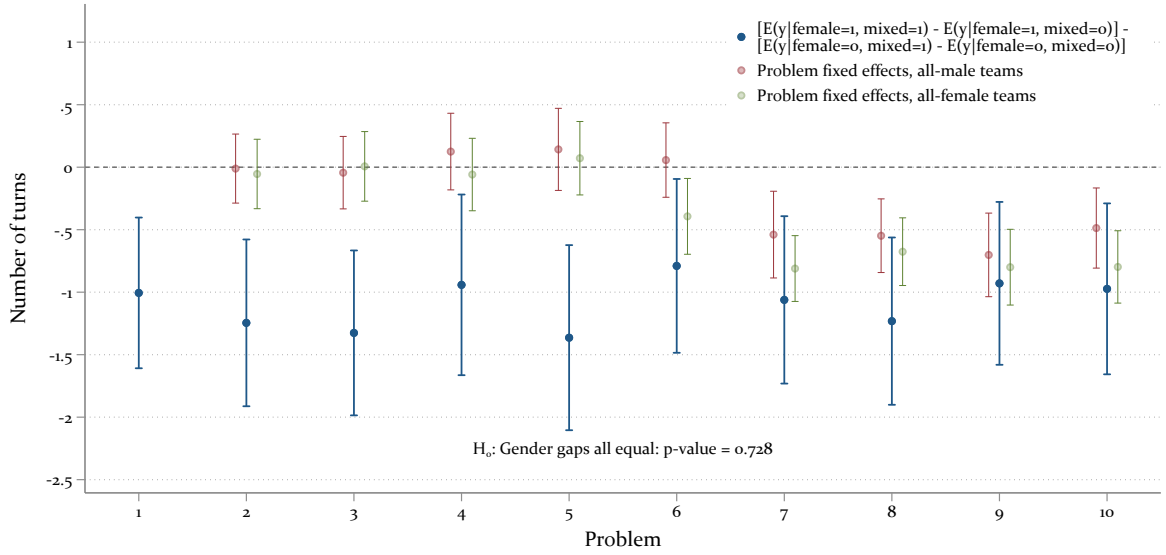
Notes: This figure shows kernel density plots for the number of turns at individual level, for subjects assigned to gender-homogenous ($N = 916$) and mixed teams ($N = 452$).

Figure B.6: Mixed Teams: Gender Composition of Subjects Ranking First and Second in Number of Words



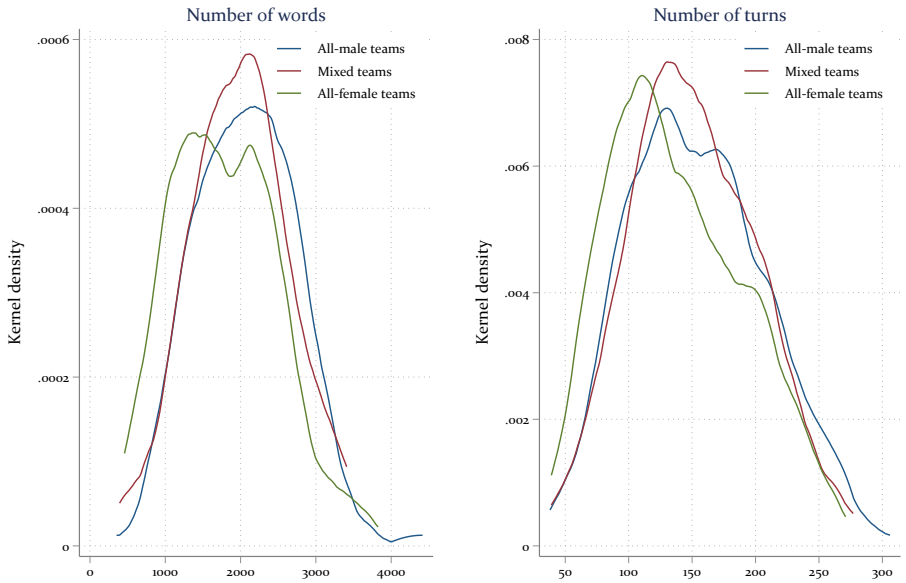
Notes: This figure displays the gender composition of subjects who rank first and second in mixed teams in terms of the number of words. The sample consists of all gender-mixed teams. The leftmost bar shows the percentage of all such teams where the females rank first and second in terms of the number of words contributed to the team's conversation. The other bars display corresponding percentages for the remaining cases: a female ranks first, a male second; a male ranks first, a female second; and males rank first and second. The sample consists of all 452 subjects assigned to gender-mixed teams.

Figure B.7: Gender Gap in Number of Turns by Problem, Individual Level



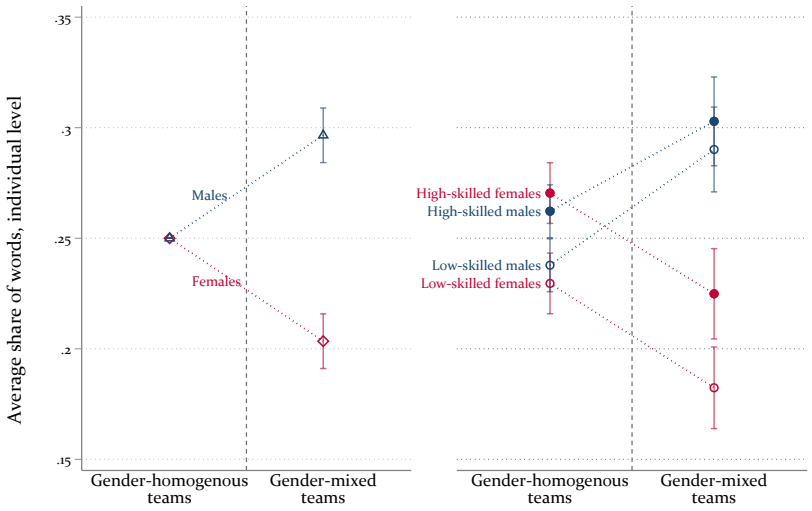
Notes: This figure is derived from an OLS regression of equation (3). The figure displays problem-specific gender gaps $\hat{\theta}_p$ for $p = 1, \ldots, 10$ (blue dots), together with 95% confidence intervals. For comparison, the figure also displays $\hat{\beta}_p$ for $p = 2, \ldots, 10$ (problem fixed effects for males in all-male teams, red dots). The problem fixed effects for females in all-female teams (green dots) are derived from an equivalent regression that uses an indicator for males (plus corresponding interactions) instead of an indicator for females. The estimations use all $1386 \times 10 = 13860$ observations.

Figure B.8: Quantity of Communication, Team Level



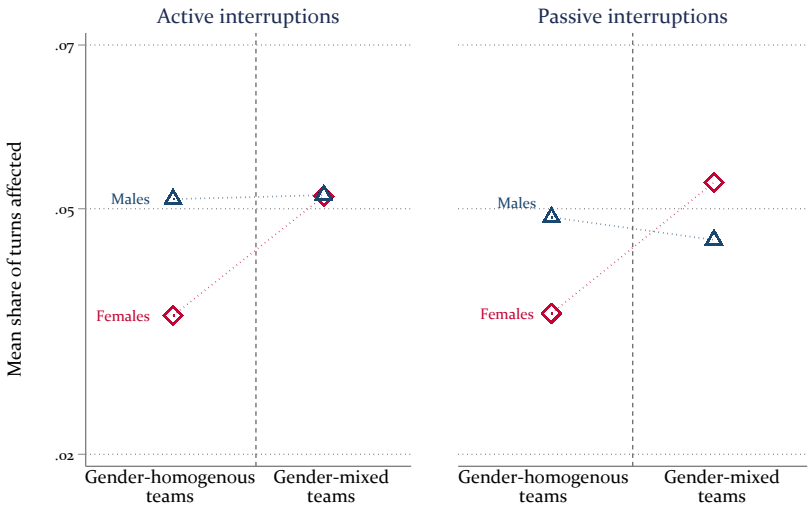
Notes: This figure shows team-level kernel density plots for the number of words and the number of turns, respectively. The sample consists of 114 all-male, 113 mixed, and 115 all-female teams.

Figure B.9: Gender Gap in Team Communication: Share of Turns



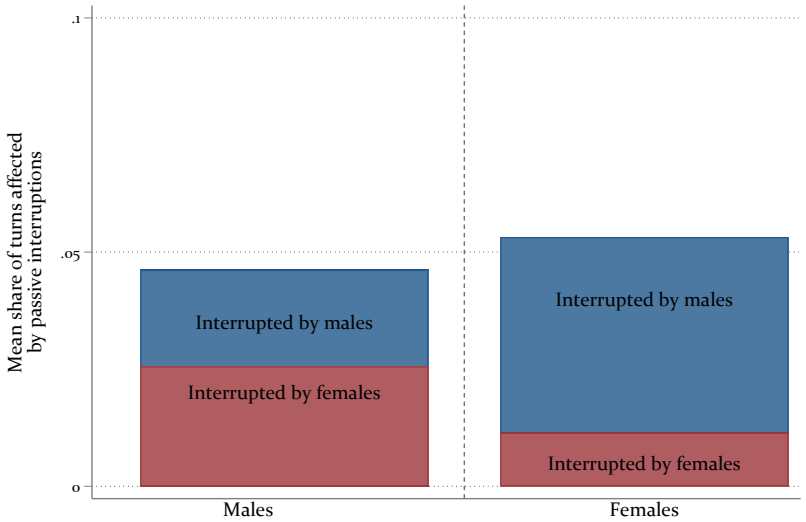
Notes: This figure displays gender gaps in team communication by team gender composition and cognitive skills. The left panel shows shares in the total number of turns at the team level spoken by female and male subjects, separately for gender-homogenous and gender-mixed teams. The right panel differentiates between subjects of above-median ("high-skilled") and below-median ("low-skilled") cognitive skills in terms of A-level GPA. The sample consists of all 1386 subjects.

Figure B.10: Active and Passive Interruptions



Notes: This figure shows the frequencies of active and passive interruptions. The sample consists of all 1386 subjects.

Figure B.11: Passive Interruptions in Mixed Teams



Notes: This figure shows the frequencies of passive interruptions in mixed teams by the subject's gender and the gender of the interrupting subject. The sample consists of all 452 subjects assigned to mixed teams.

C Communication Measures in Python

We extract various communication measures from both audio files and written transcripts. The transcripts include information on the speaker and timestamps for the beginning of each turn. Additionally, the transcripts also mark interruptions. We transcribed the audio files separately by team and problem. When lemmatizing the transcripts,[48] we manually added lemmas for German words that were missing in the respective database. Each lemma was assigned a team, a problem, a speaker, and a turn. For the lexical sentiment analysis, we also assigned it to a sentence.

To derive the number of *words*, we counted all words in the transcripts except for filler words such as "oh" or "hm". For the number of *turns*, we counted all turns consisting of at least 3 words. To measure *interruptions*, we counted the coded interruptions if a turn of at least 3 words interrupted another turn of at least 3 words. For *topic words*, we counted the words defined as topic words among all lemmatized words. For the *lexical sentiment analysis*, we counted sentiment words in the non-lemmatized words, and if a sentiment word was part of a negated sentence, its value was multiplied by -1.

To derive measures of *speaking time* and *sentiment* from the audio files, we used the transcripts' timestamps indicating the beginning of each turn for dividing the audio into snippets. We then removed from the snippets periods of silence exceeding a length of two seconds.[49] We then transferred the snippets to 16 kHz. To calculate *total speaking time*, we aggregated the lengths of the silence-reduced snippets at the speaker and team level.

For the analysis of *sentiment*, we trained our models on the emoDB database (Burkhardt et al., 2005), which includes German-spoken sentences in different emotions, all reduced to 16 kHz. We consider the emotions "happy", "sad", and "neutral", and further divided the data by gender to generate two distinct models. Before training the models, we reduced the dimensions of the audio files by computing the Mel-Frequency Cepstral Coefficients (MFCCs) and keeping 13 coefficients for the further steps.[50] We then created an LSTM model with two additional layers and a softmax layer.[51] We allocated 70% of the selected data for training and 30% for testing, resulting in a male model with an overall accuracy of 92.59%. It achieved 100% accuracy in recognizing the emotions "happy" and "neutral", and 75% accuracy in identifying the emotion "sad". The female model achieved an overall accuracy of 97.22% (100% accuracy in

[48] See the package SpaCy, `https://spacy.io/`.

[49] For this step, we used the package pydub, `https://github.com/jiaaro/pydub/`.

[50] We reduced the audio files using the package tensorflow, `https://www.tensorflow.org/`.

[51] We used the package Keras, `https://keras.io/`.

recognizing “sad” and “neutral”, and 92% accuracy in identifying “happy”). Our model was run on a system equipped with 8 Premium Intel CPUs.

Our trained model was then used to predict the emotions in the snippets, which were also transformed into the MFCCs representation. At the snippet level, the output consists of a weight for each of the three emotions, with the weights for each snippet adding up to 1. We then derive our *sentiment* measures by averaging the weights over a speaker’s turns, weighted by the turns’ length.

D Lexical Sentiment Score

In the pre-analysis plan, we committed to running regressions at the team level using a lexical sentiment score following Remus et al. (2010). This regression was meant to capture differences in the sentiment of the team conversation between teams of different gender compositions. The lexical approach rests on the idea of comparing the individual words that subjects used in the team conversation with predefined lists of words, w, bearing negative and positive sentiment weights $s_w \in [-1;1]$. When a sentence was negated (or a part of it), we used the additive inverse of the original weight of the negated part. The sentiment score at the team level is then derived by summing up the weights of all words spoken by a team and dividing by the number of sentiment words.

When analyzing the transcriptions of the audio files capturing the teams' conversations during the team task, we became aware that the usage of a sizeable share of the words carrying a sentiment weight seemed to be triggered by the fact that the team task was designed as a single-choice decision problem. To demonstrate this issue that was unforeseen by us when pre-specifying the data analysis, Table D.1 reports the 15 words carrying the highest polarity weights, separately for positive and negative sentiment words. The analysis is based on all appearances of sentiment words across the conversations of all 342 teams. A word's polarity weight measures the share of the overall (positive or negative) polarity of verbal communication across all teams determined by the usage of this word and is derived by first calculating a word's aggregate polarity by multiplying the overall number of appearances of the word in the data with the absolute value of its polarity and then dividing this aggregate polarity by the sum of aggregate polarities over all the positive (negative) sentiment words.

The left panel of the table shows that, out of 1165 different positive sentiment words used by all teams, the 15 most influential words determine 69.4 percent of the aggregate positive polarity of team conversation. Similarly, the right panel of the table demonstrates that, out of the 1050 different negative sentiment words, the 15 most influential words determine 73.5 percent of the aggregate negative polarity. The frequent usage of several of the listed words is likely triggered by the fact that the team task was a single-choice task. For instance, the teams often used the word "exclude" (or versions thereof) when discussing the likelihood of certain statements being true. Similarly, the subjects often used "good", "better", "bad", "wrong", "sure", "NOT sure", and "unsure" when assessing their options to answer a single-choice problem. The usage of "illness" was likely triggered by the fact that one of the blocks of single-choice problems referred to a business case featuring a pharmaceutical company.

Table D.1 suggests that both the positive and the negative lexical sentiment scores are largely determined by the usage of words that reflect the type of the team task rather than the true sentiment of the team conversation. We, therefore, decided to deviate from the pre-analysis plan in terms of the measurement of team sentiment and use vocal features following Hu and Ma (2021) instead of lexical sentiment scores.

For completeness, Table D.2 reports the pre-specified regression based on the lexical sentiment score. In line with the notion that the lexical score is dominated by words triggered by our design, the team gender composition does not affect the lexical score.

Tables 8 in the paper and A.23 in this Online Appendix report the results for sentiment based on vocal features. Online Appendix Section C provides further details.

Table D.1: Composition of Lexical Sentiment Score

Aggregate weight of words with positive polarity		Aggregate weight of words with negative polarity	
good	0.306	excluded	0.220
better	0.077	bad	0.139
big	0.060	wrong	0.122
NOT bad	0.034	slight	0.050
important	0.031	NOT sure	0.042
NOT excluded	0.025	NOT helping	0.028
perfect	0.024	illness	0.021
sure	0.021	little	0.018
like	0.021	unsure	0.018
super	0.018	NOT good	0.018
helping	0.019	end	0.014
fast	0.015	dependence	0.015
growing	0.015	stupid	0.011
convinced	0.015	problem	0.011
next	0.015	falling	0.009
Total	0.695	Total	0.736

Notes: This table is based on all sentiment words spoken across all 342 teams and shows the words carrying the largest polarity weights, separately for words with positive and negative polarity. A word's polarity weight measures the share of the overall (positive or negative) polarity of verbal communication across all teams determined by the usage of this word and is calculated as follows. We first calculate a word's aggregate polarity by multiplying the overall number of appearances of the word in the data with the absolute value of its polarity. We then derive a word's polarity weight in the positive (negative) sentiment score by dividing its aggregate polarity by the sum of aggregate polarities over all the positive (negative) sentiment words.

Table D.2: Effects on Lexical Sentiment Score

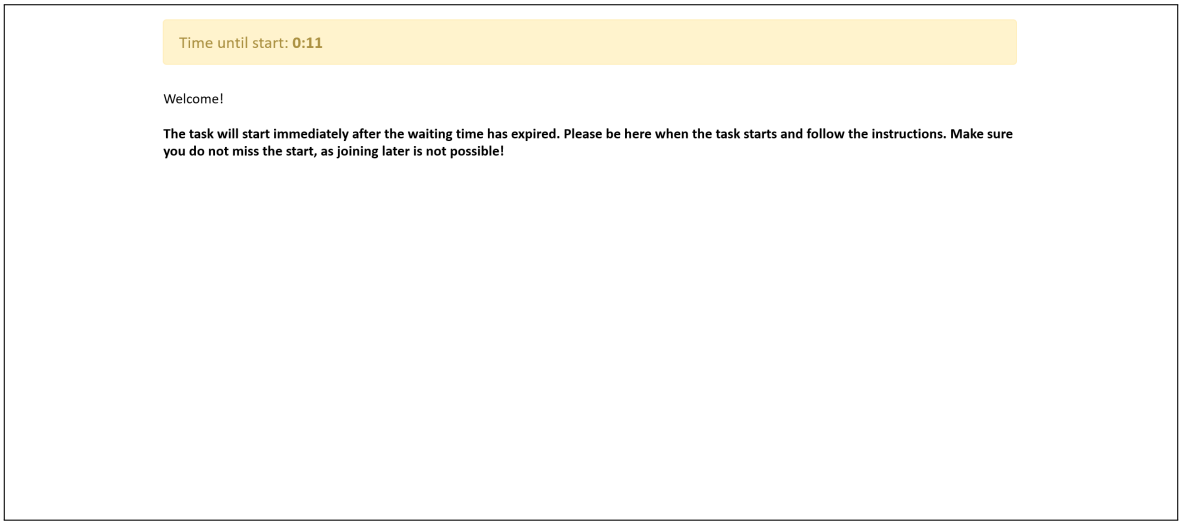
	Lexical sentiment score
Gender-mixed team (β_1)	-0.006
	(0.004)
All-female team (β_2)	-0.008
	(0.005)
N. of obs.	342
Mean dep. var. all-male	-0.01
Team-level controls	Yes
$\beta_1 = \beta_2$ (p-value)	0.626
$\beta_1 = 0$ (p-value MHT)	0.180
$\beta_2 = 0$ (p-value MHT)	0.199

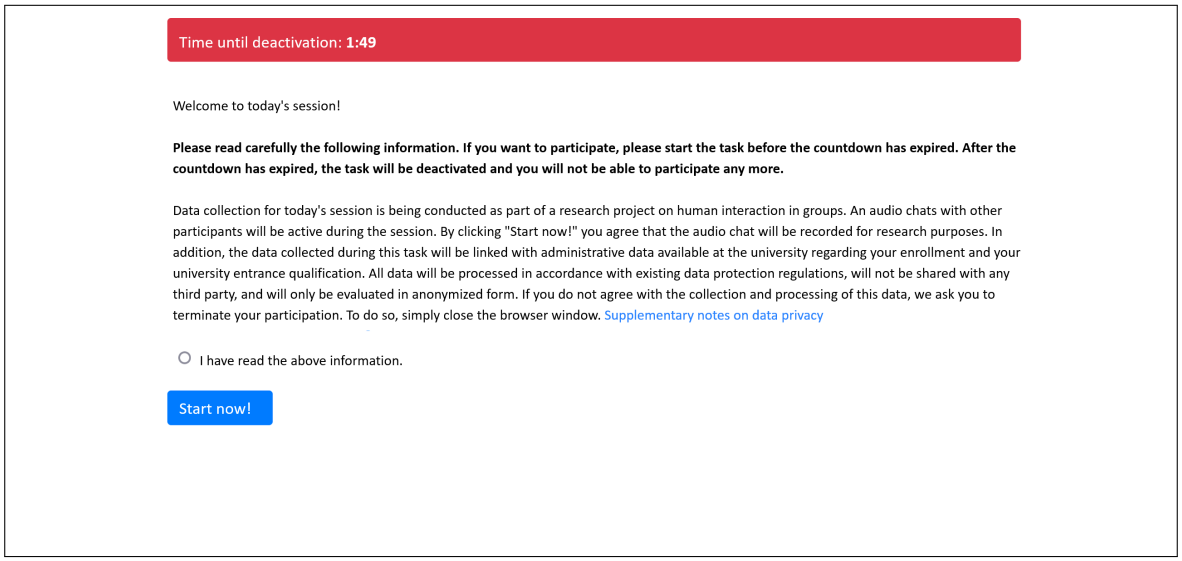
Notes: This table shows a team-level OLS regression using a lexical sentiment score as dependent variable (the sentiment-related outcome we committed to use in the pre-analysis plan). The regression controls for team averages of A-level GPA and age, maximum and minimum A-level GPA, maximum and minimum age, the share of team members with an A-level degree obtained from top-tier high school type, the share of team members with foreign nationality, the share of team members studying at Master level, a series of variables capturing the shares of team members studying in one of the main study fields (arts and humanities, engineering, natural sciences, economics/business administration), and an indicator for teams where some members were silent during the team task. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values adjusted for multiple hypothesis testing (MHT, two hypotheses included) follow Barsbai et al. (2020).

E Experimental Instructions

This section shows screenshots of stage 1 and stage 2 of the experiment (translated from German). Screenshots are in chronological order. Headings refer to Appendix Figure B.1 showing the timeline of the design.

Stage 1: Instructions and Matching





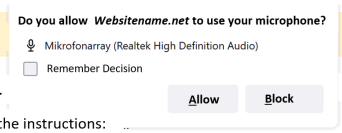
Remaining time: **1:01**

Microphone test

You will need a microphone to participate in this session.
You can only participate if you perform the microphone test now.
Please click on "Start microphone test" to test your microphone.

Start microphone test

Remaining

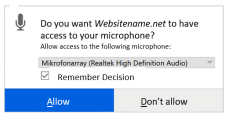


You still have **1** to successfully complete the microphone test.

Please follow the instructions:

1. **Enable microphone use**

   Click "**Allow**" or "Permit" when you receive one of the following messages.



oder

2. **Unmute**

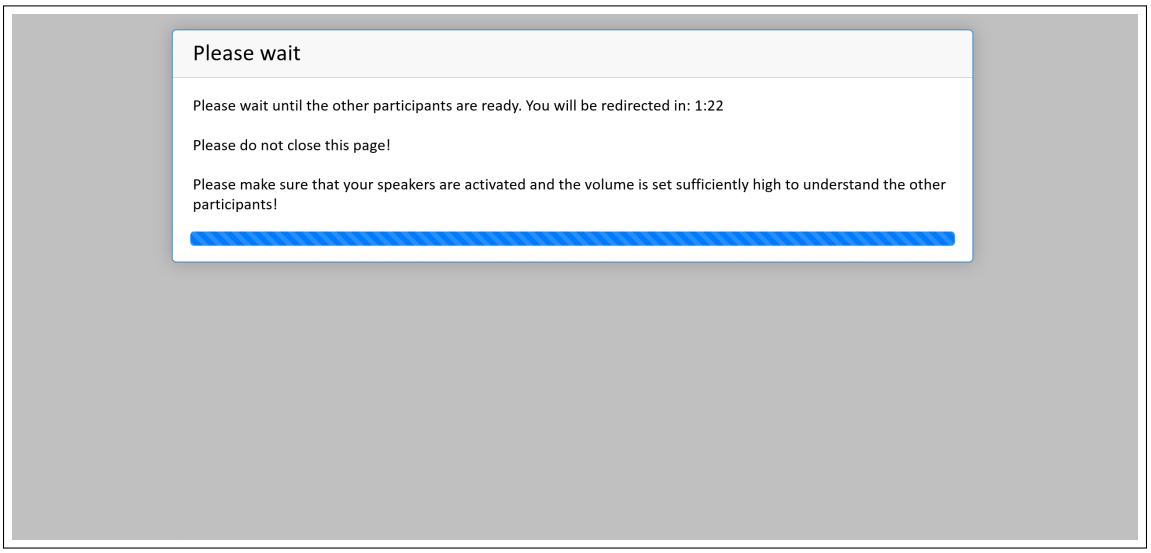
   Your microphone is muted when the following icon is displayed in the audio chat window on the right edge of your screen:  **Click on the icon to unmute.**

   You have been successfully unmuted when the following icon is displayed:

Remaining time on this page: **0:10**

You successfully passed the microphone test.

Please make sure that your speakers are activated and the volume is set sufficiently high to understand the other participants.



Time remaining until automatic redirecting: **0:30**

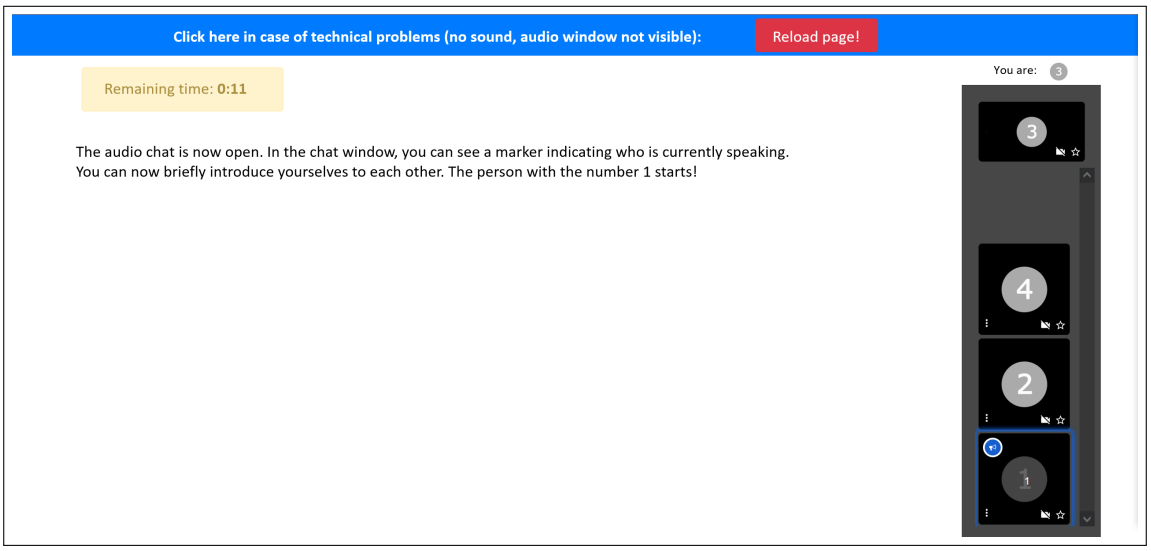
Instructions Part 1

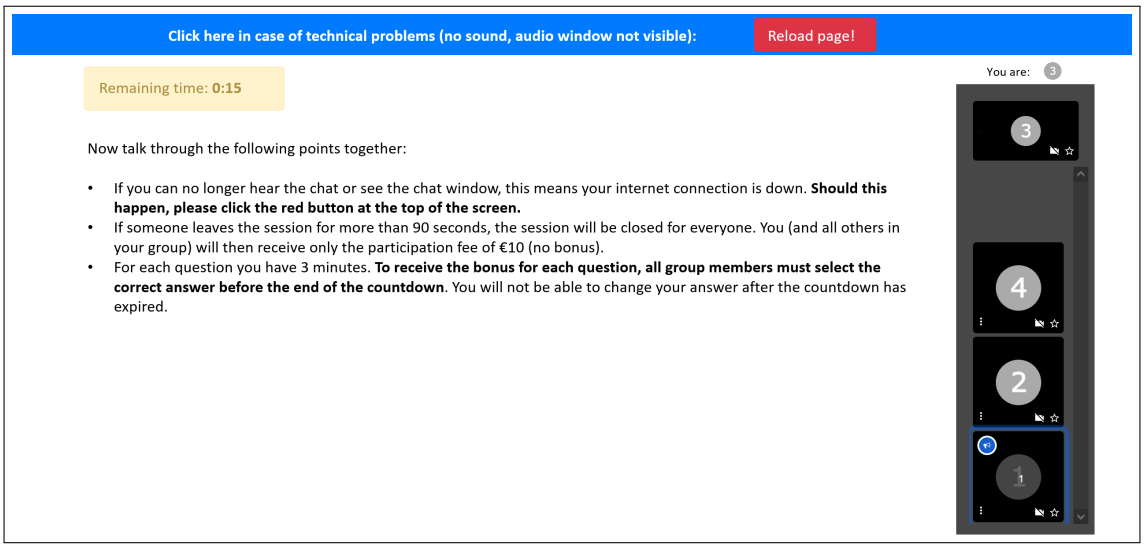
- Today's session is divided into two parts. For participation in the first part of the session you will receive a fixed payment of €10. In addition, you can earn additional money in the first and second part of the session. However, you will receive these variable payoffs only if you provide complete information in the corresponding sections. The first part of the session starts after the automatic redirection.
- You will be matched with 3 other randomly selected participants.
- You can communicate with the other participants via an audio chat.
- You will work with your group on a total of 10 multiple-choice questions. Each question has exactly one correct answer.
- Only if all 4 group members choose the correct answer, you (and every other member of your group) will receive a bonus of €1 for this question. If someone in your group does not choose the correct answer, no one in the group will receive a bonus.
- With 10 questions in total, you can earn a bonus of up to €10.

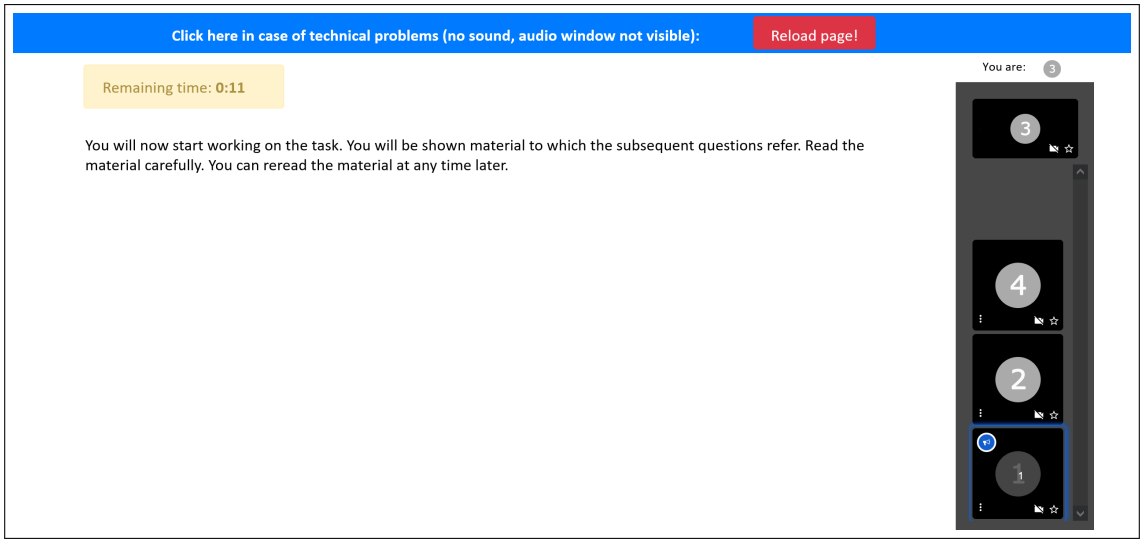
Conference starts in: **0:16**

Instructions Part 1 (continued)

- After redirecting, you will be connected to the other group members.
- All members remain anonymous. You can address each other with the numbers visible in the audio chat window.
- First check if you can communicate with the other group members.
- Then go through the points displayed on the page together.

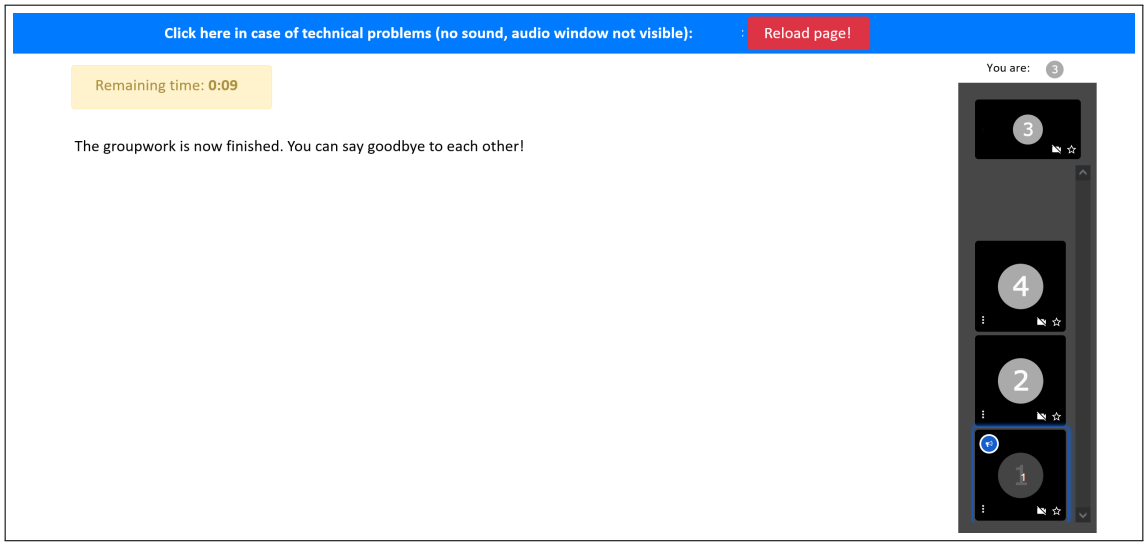

Click here in case of technical problems (no sound, audio window not visible):
Reload page!
Remaining time: 0:11
The audio chat is now open. In the chat window, you can see a marker indicating who is currently speaking.
You can now briefly introduce yourselves to each other. The person with the number 1 starts!
You are: 3
3
4
2
1


Click here in case of technical problems (no sound, audio window not visible):
Reload page!
Remaining time: 0:15
Now talk through the following points together:
• If you can no longer hear the chat or see the chat window, this means your internet connection is down. Should this happen, please click the red button at the top of the screen.
• If someone leaves the session for more than 90 seconds, the session will be closed for everyone. You (and all others in your group) will then receive only the participation fee of €10 (no bonus).
• For each question you have 3 minutes. To receive the bonus for each question, all group members must select the correct answer before the end of the countdown. You will not be able to change your answer after the countdown has expired.
You are: 3
3
4
2
1

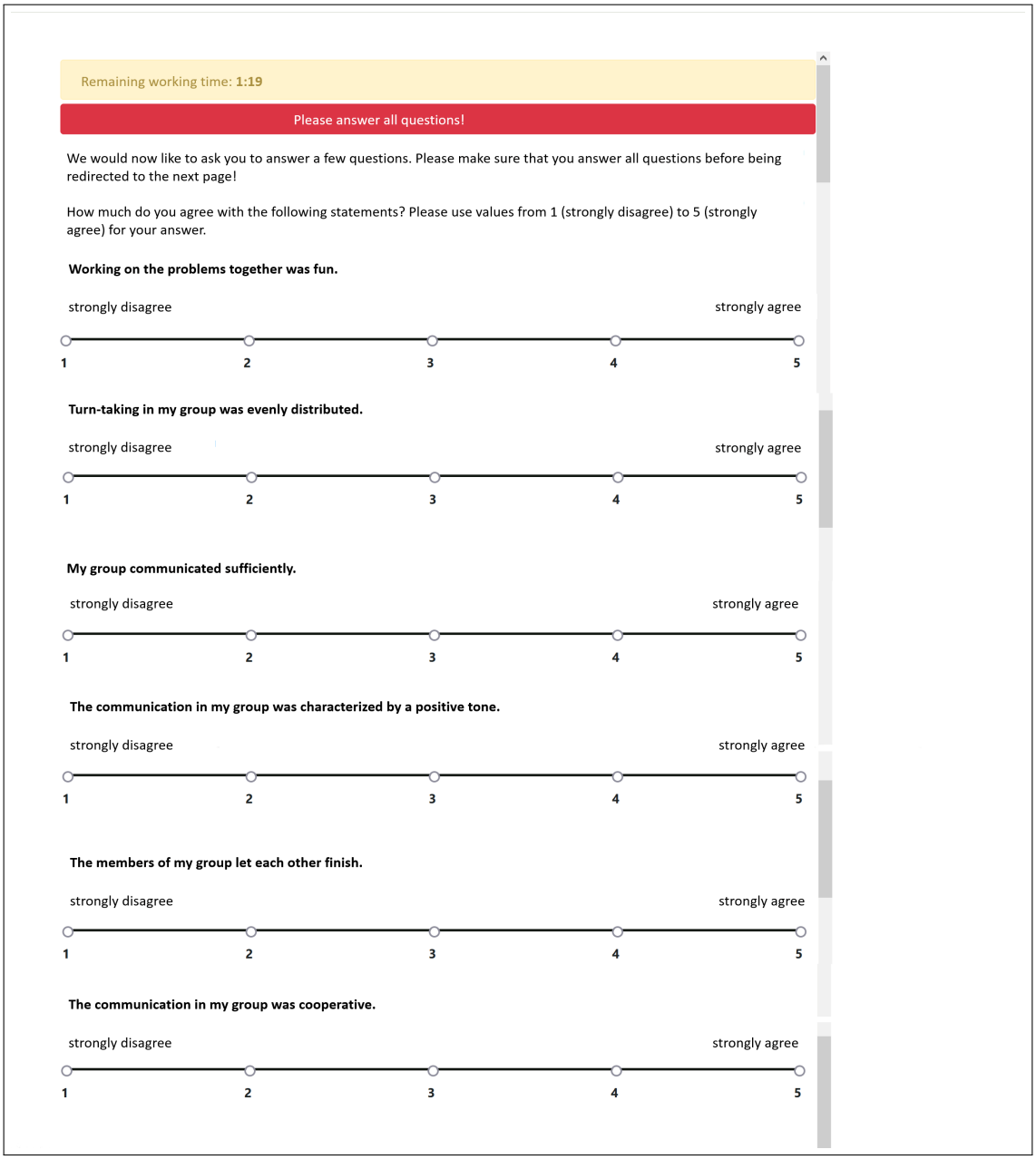

Click here in case of technical problems (no sound, audio window not visible):
Reload page!
Remaining time: 0:11
You will now start working on the task. You will be shown material to which the subsequent questions refer. Read the material carefully. You can reread the material at any time later.
You are: 3
3
4
2
1

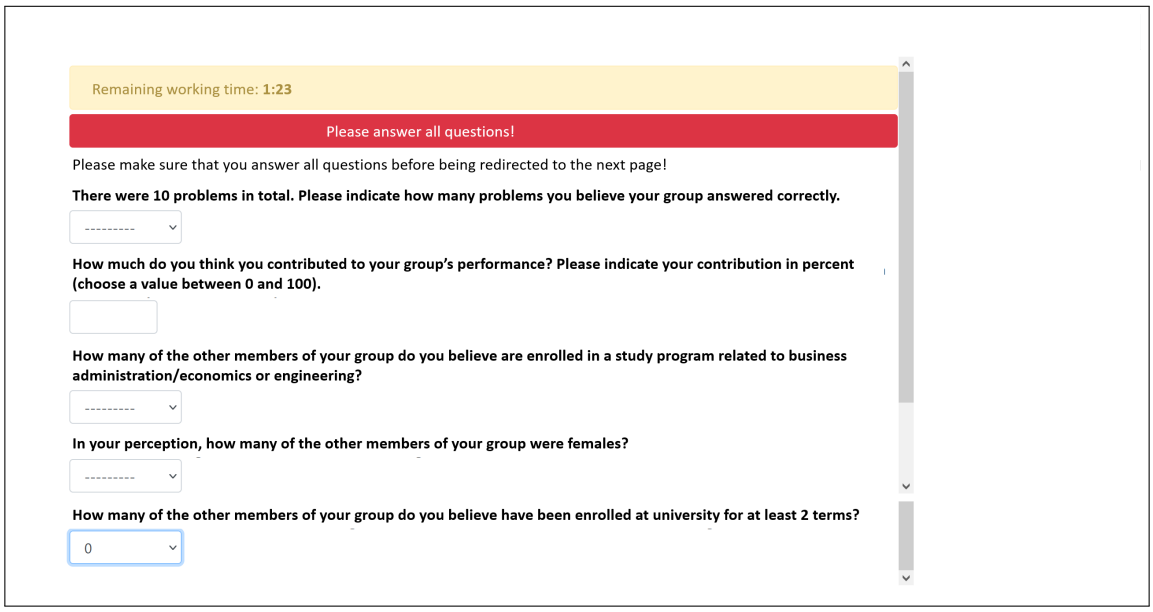
Stage 1: Team Task

At this point, the subjects started working on the real effort task (30 minutes plus reading time). While working on the 10 different problems, the subjects could study instructions and information material by opening and closing tabs. Here, we show only the stage-1 farewell screen. Appendix Section F displays two sample screenshots of the team task.

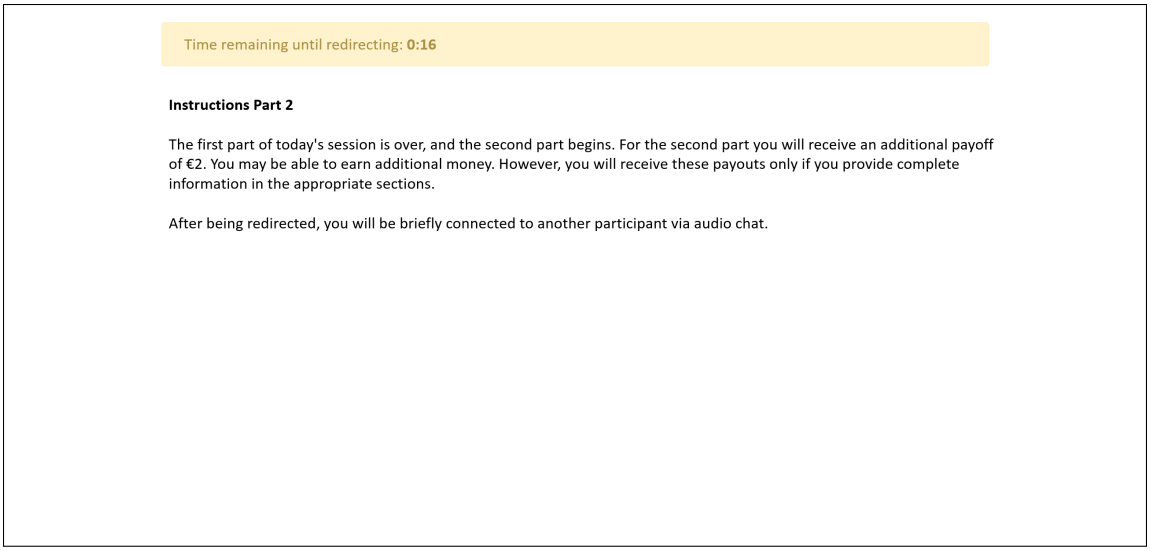


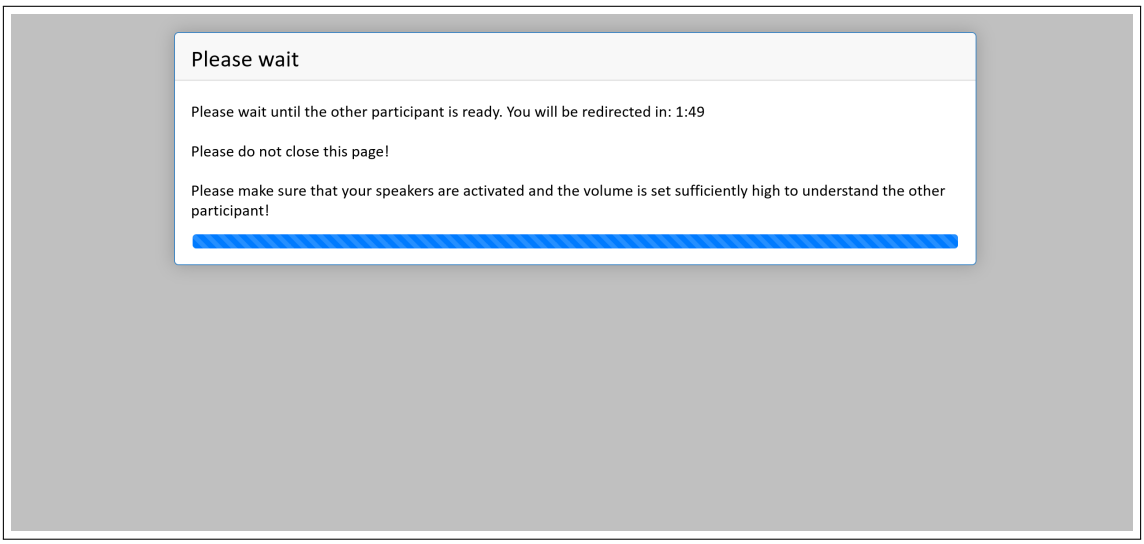
Stage 1: Survey



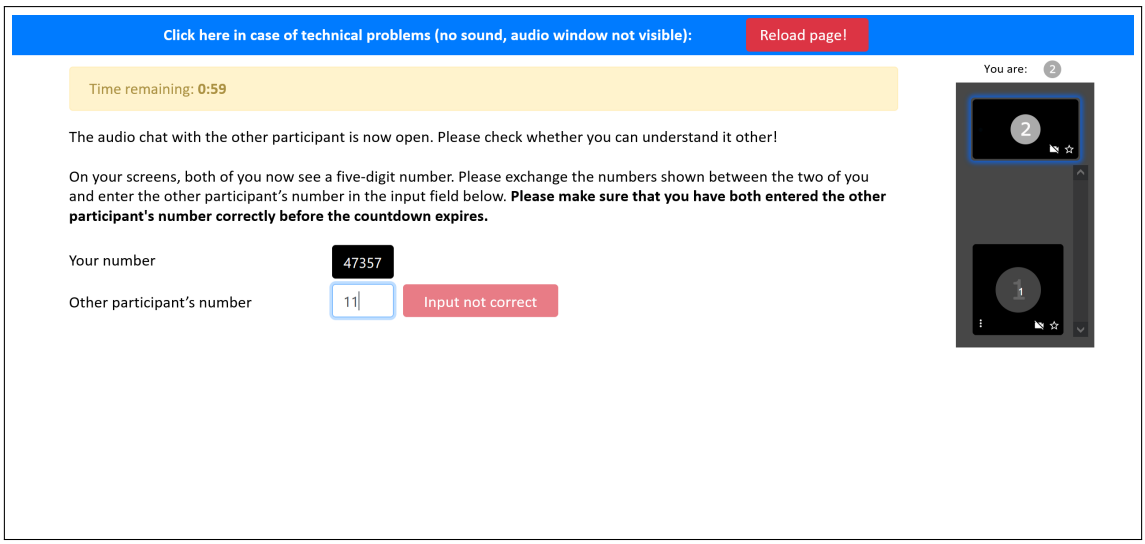


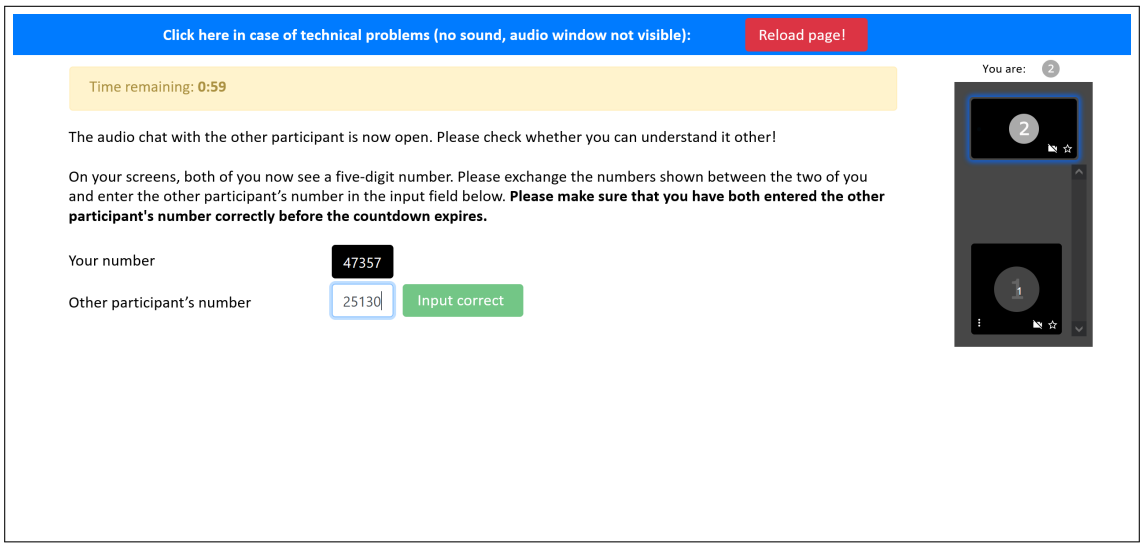
Stage 2: Instructions and Matching



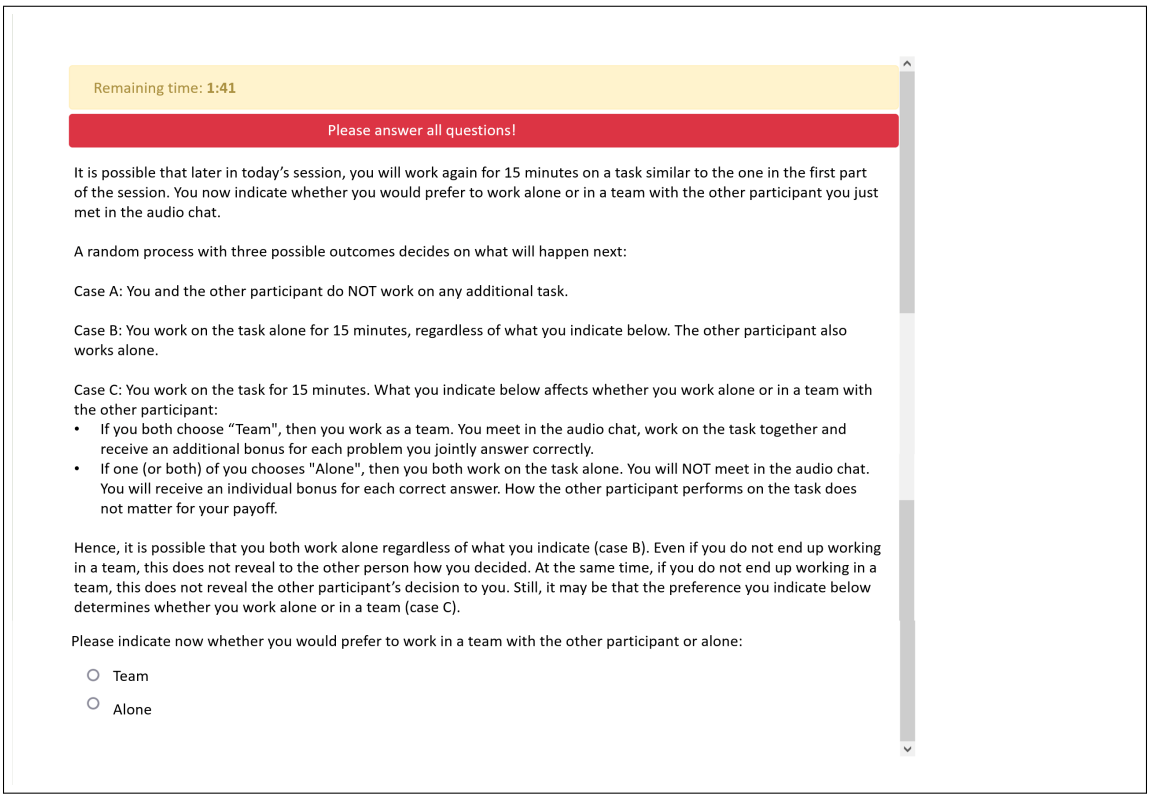


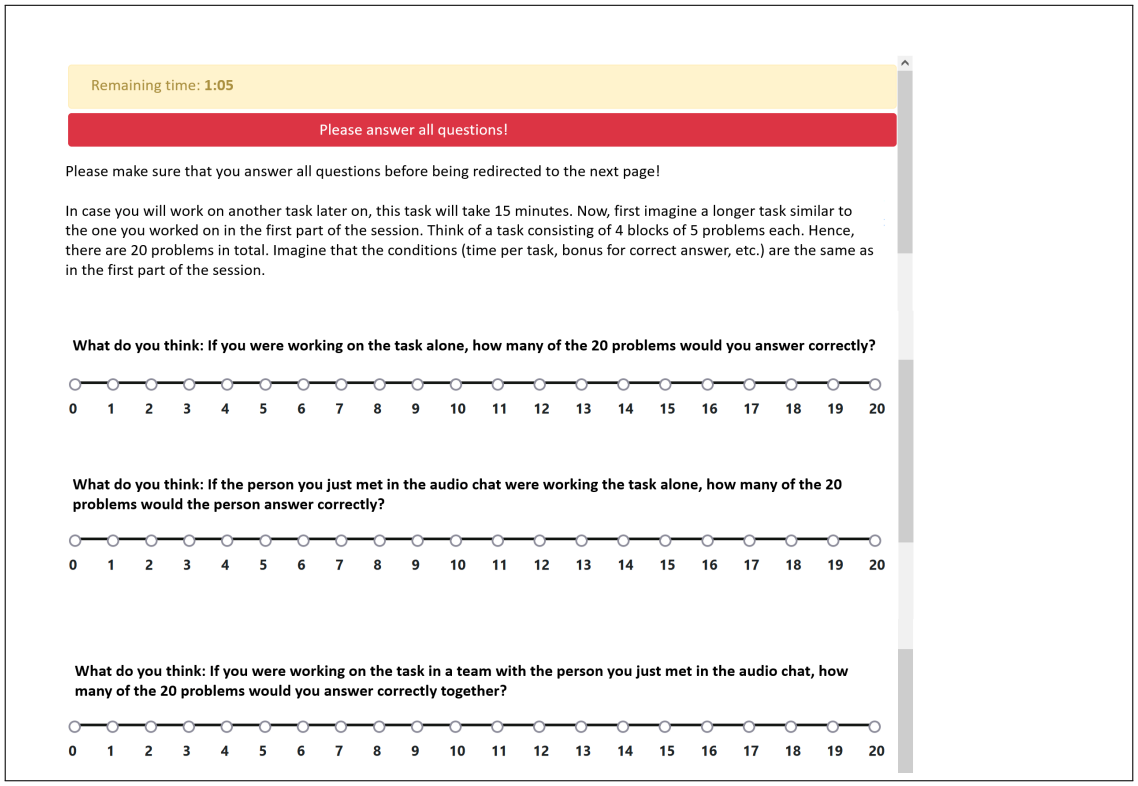
Stage 2: Exchange of Keys

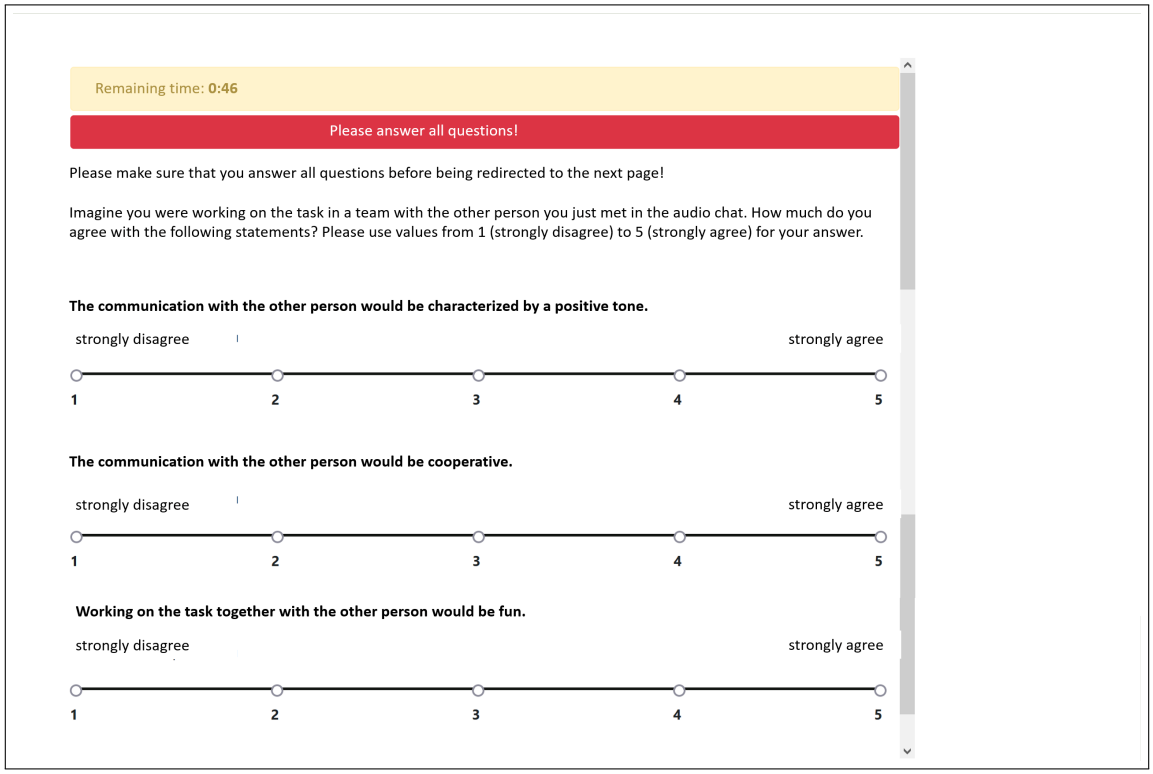




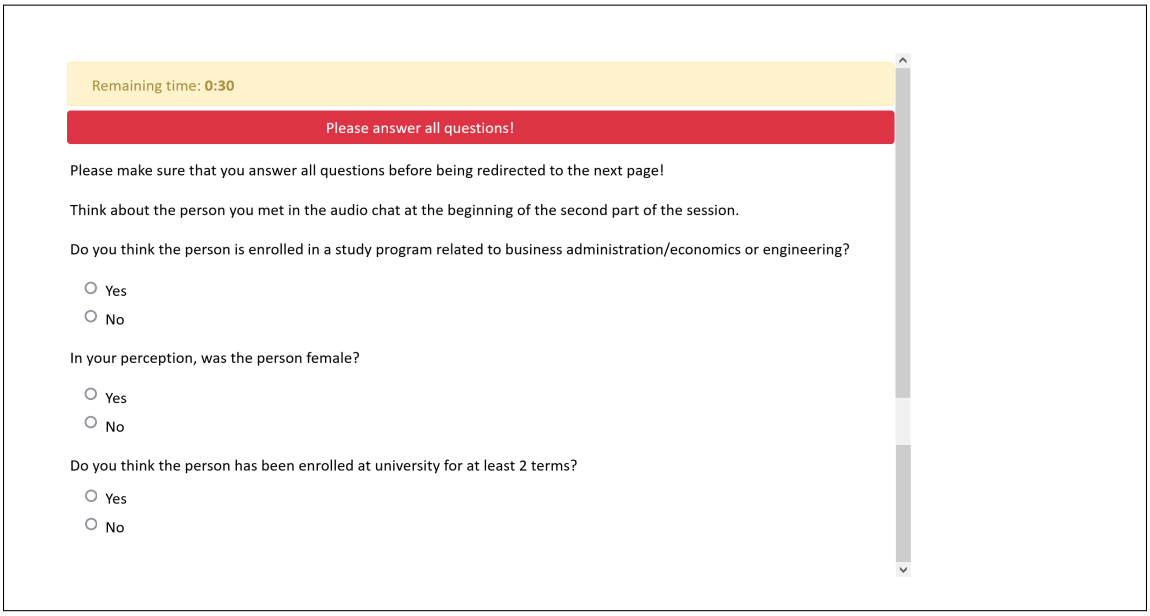
Stage 2: Elicitation of Preferences and Beliefs




Remaining time: 1:05
Please answer all questions!
Please make sure that you answer all questions before being redirected to the next page!
In case you will work on another task later on, this task will take 15 minutes. Now, first imagine a longer task similar to the one you worked on in the first part of the session. Think of a task consisting of 4 blocks of 5 problems each. Hence, there are 20 problems in total. Imagine that the conditions (time per task, bonus for correct answer, etc.) are the same as in the first part of the session.
What do you think: If you were working on the task alone, how many of the 20 problems would you answer correctly?
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
What do you think: If the person you just met in the audio chat were working the task alone, how many of the 20 problems would the person answer correctly?
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
What do you think: If you were working on the task in a team with the person you just met in the audio chat, how many of the 20 problems would you answer correctly together?
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20


Remaining time: 0:46
Please answer all questions!
Please make sure that you answer all questions before being redirected to the next page!
Imagine you were working on the task in a team with the other person you just met in the audio chat. How much do you agree with the following statements? Please use values from 1 (strongly disagree) to 5 (strongly agree) for your answer.
The communication with the other person would be characterized by a positive tone.
strongly disagree
strongly agree
1 2 3 4 5
The communication with the other person would be cooperative.
strongly disagree
strongly agree
1 2 3 4 5
Working on the task together with the other person would be fun.
strongly disagree
strongly agree
1 2 3 4 5

Stage 2: Survey



Remaining time: **0:30**

Please answer all questions!

Please make sure that you answer all questions before being redirected to the next page!

Think about the person you met in the audio chat at the beginning of the second part of the session.

Do you think the person is enrolled in a study program related to business administration/economics or engineering?

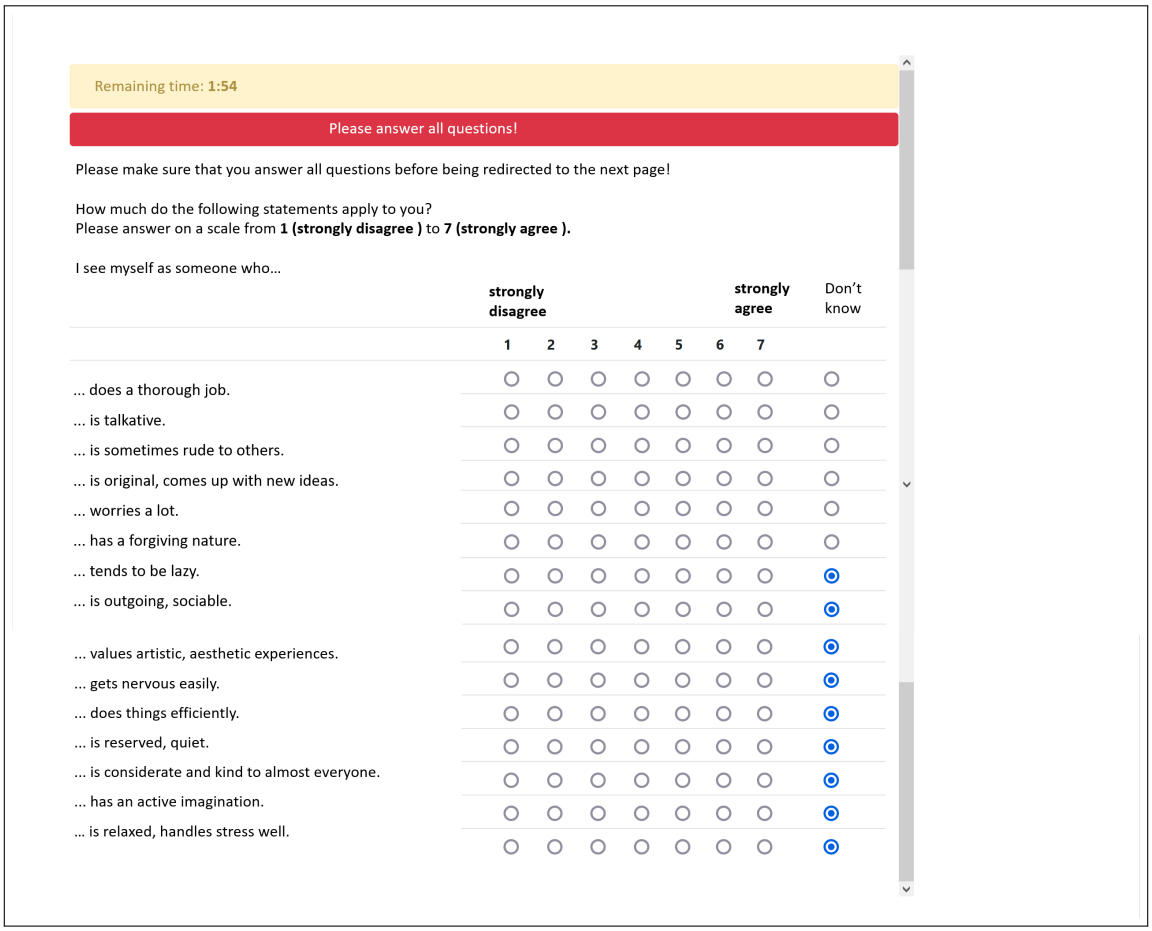
- Yes
- No

In your perception, was the person female?

- Yes
- No

Do you think the person has been enrolled at university for at least 2 terms?

- Yes
- No



Remaining time: **1:54**

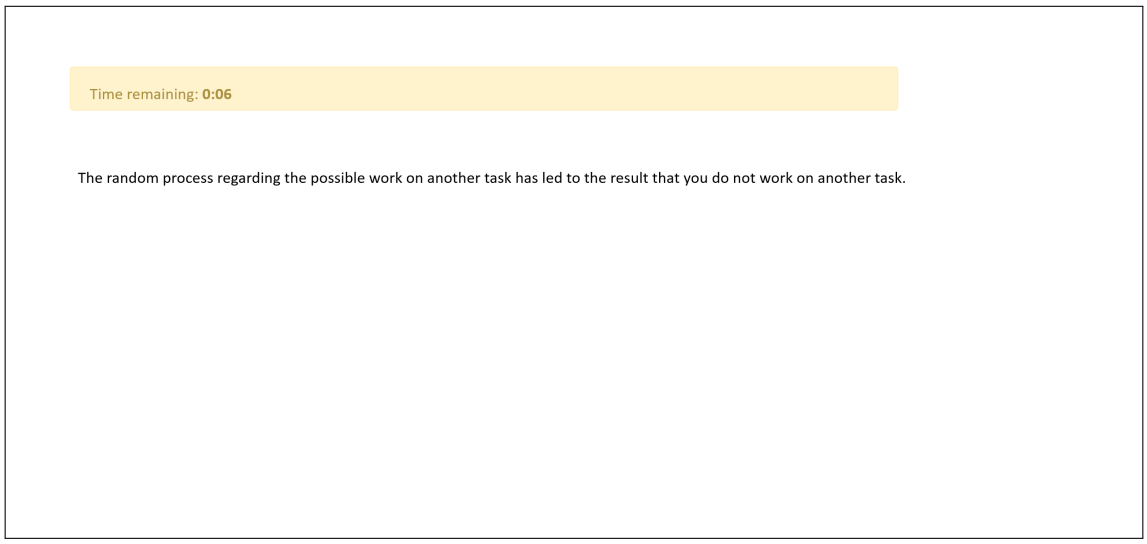
Please answer all questions!

Please make sure that you answer all questions before being redirected to the next page!

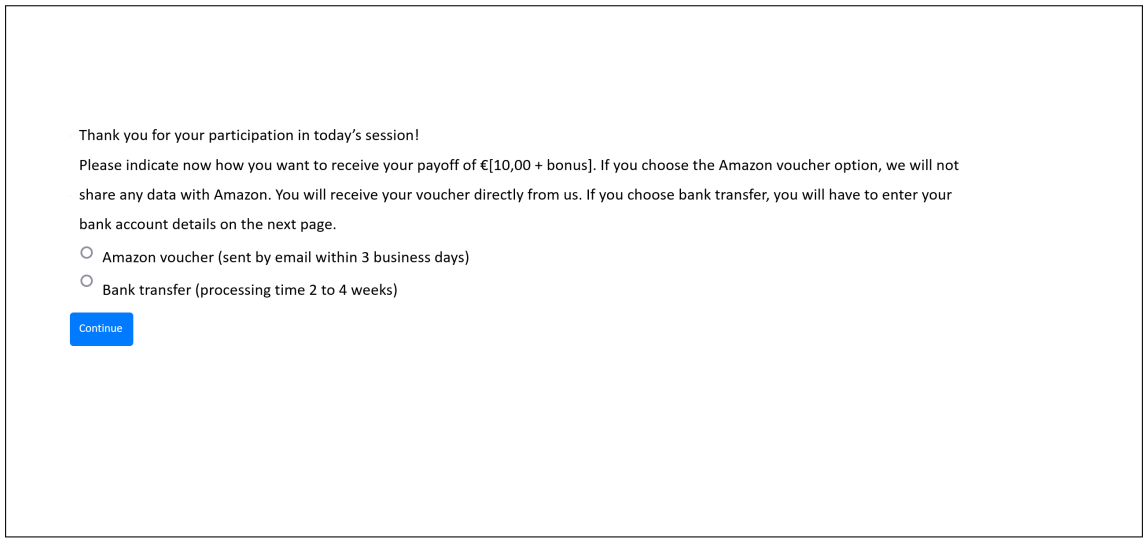
How much do the following statements apply to you?
Please answer on a scale from **1 (strongly disagree)** to **7 (strongly agree).**

I see myself as someone who...

	strongly disagree						**strongly agree**	Don't know
	1	**2**	**3**	**4**	**5**	**6**	**7**	
... does a thorough job.	○	○	○	○	○	○	○	○
... is talkative.	○	○	○	○	○	○	○	○
... is sometimes rude to others.	○	○	○	○	○	○	○	○
... is original, comes up with new ideas.	○	○	○	○	○	○	○	○
... worries a lot.	○	○	○	○	○	○	○	○
... has a forgiving nature.	○	○	○	○	○	○	○	○
... tends to be lazy.	○	○	○	○	○	○	○	◉
... is outgoing, sociable.	○	○	○	○	○	○	○	◉
... values artistic, aesthetic experiences.	○	○	○	○	○	○	○	◉
... gets nervous easily.	○	○	○	○	○	○	○	◉
... does things efficiently.	○	○	○	○	○	○	○	◉
... is reserved, quiet.	○	○	○	○	○	○	○	◉
... is considerate and kind to almost everyone.	○	○	○	○	○	○	○	◉
... has an active imagination.	○	○	○	○	○	○	○	◉
... is relaxed, handles stress well.	○	○	○	○	○	○	○	◉



Payoff Screen and Selection of Payment Method

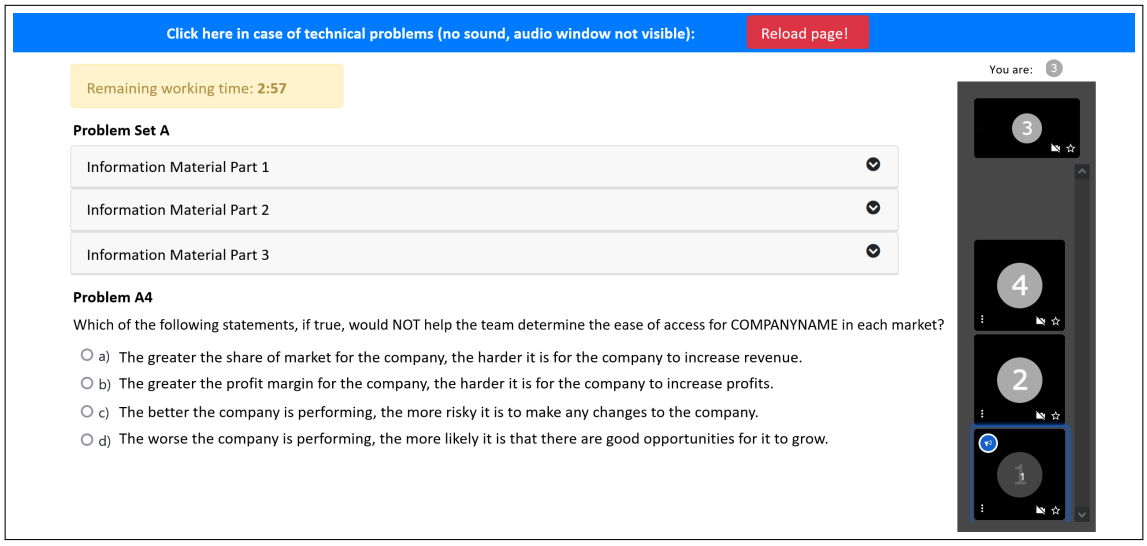


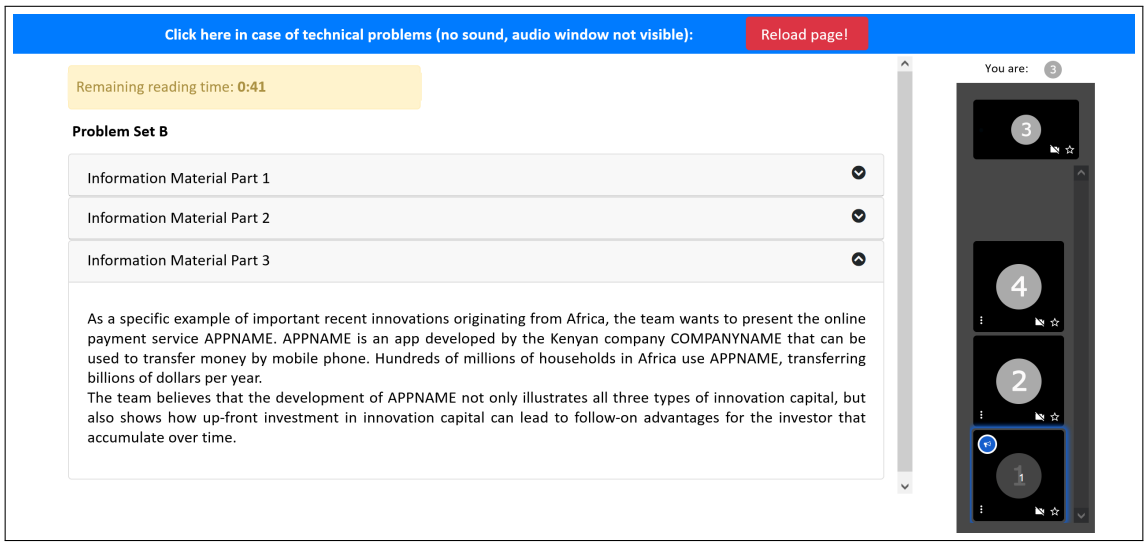
F Team Task

The task consisted of a series of 10 single-choice problems, grouped into two problem sets. Each set of problems referred to a business case that was described using extensive information material. The first business case was concerned with a hypothetical firm. The problems referred to issues related to the firm's sales and profits, as well as investments and market access in different world regions. The second business case dealt with economic development in Africa, with a focus on different forms of capital, investment and innovation.

Whenever new information material was introduced, teams were given extra time for studying the material. When working on the problems, the team members could go back to this material at all times by opening and closing tabs. Once the three minutes for a given problem had elapsed, the subjects could no longer access this problem, and answers to this problem could no longer be changed. In order to earn a bonus for a given problem, all four members of a given team had to mark the correct statement on their screen. Coordination among team members was only possible via the audio chat, which was open throughout the team task.

In the following, we document two sample screenshots of the team task.




Click here in case of technical problems (no sound, audio window not visible):
Reload page!
Remaining reading time: 0:41
Problem Set B
Information Material Part 1
Information Material Part 2
Information Material Part 3
As a specific example of important recent innovations originating from Africa, the team wants to present the online payment service APPNAME. APPNAME is an app developed by the Kenyan company COMPANYNAME that can be used to transfer money by mobile phone. Hundreds of millions of households in Africa use APPNAME, transferring billions of dollars per year.
The team believes that the development of APPNAME not only illustrates all three types of innovation capital, but also shows how up-front investment in innovation capital can lead to follow-on advantages for the investor that accumulate over time.
You are: 3
3
4
2
1