Supplemental Material

Method

Participants

Twenty-six groups had four people (two groups with four women, six groups with three women, thirteen groups with two women, and five groups with one woman) and five groups had three people (two groups with three women, two groups with two women, and one group with one woman).

Participants said they did not know a particular groupmate at all before the study 62.3% of the time. When participants did know a groupmate before the study, they reported not knowing the groupmate very well ($M = 3.47$, $SD = 1.68$) on a scale of 1 (*not very well*) to 7 (*extremely well*). Of those who reported knowing each other, nine reported knowing their partner a "6" on the scale, and six reported a "7". Thus, a very small percentage of the sample (4.39%) considered each other close acquaintances or friends. Among those who did know each other, the majority had taken a class together in the past (we asked how they knew each other, and this was the most frequent response).

Table S1. *Home countries of participants.*

Country	Frequency	Gender Inequality Index	Country	Frequency	Gender Inequality Index	Country	Frequency	Gender Inequality Index
Afghanistan	1	0.59	Germany	3	0.084	New Zealand	1	0.137
Albania	1	0.236	Ghana	2	0.543	Norway	1	0.047
Australia	5	0.104	Greece	3	0.124	Pakistan	4	0.549
Austria	1	0.078	Hong Kong	1	NA	Peru	1	0.383
Bosnia and Herzegovina	2	0.167	Hungary	2	0.27	Philippines	1	0.427
Brazil	1	0.404	India	4	0.522	Poland	1	0.128
Bulgaria	1	0.22	Indonesia	2	0.453	Romania	1	0.316
China	5	0.164	Italy	1	0.08	Russia	4	0.262
Colombia	2	0.411	Japan	3	0.102	Rwanda	1	0.412
Costa Rica	1	0.303	Jordan	1	0.471	Slovakia	1	0.187
Cote d'Ivore	1	0.658	Kazakhstan	1	0.204	Slovenia	1	0.053
Croatia	1	0.128	Kenya	1	0.548	South Korea	5	0.06
Czech Republic	2	0.133	Latvia	3	0.208	Sri Lanka	3	0.381
Denmark	1	0.041	Lithuania	1	0.129	Taiwan	2	NA
Egypt	2	0.452	Malaysia	1	0.286	Ukraine	3	0.289
El Salvador	1	0.398	Mexico	2	0.346	United Arab Emirates	6	0.138
England	2	0.128	Moldova	1	0.232	United States	10	0.202
Ethiopia	3	0.512	Nepal	6	0.488	Vietnam	1	0.318
Finland	1	0.055						

Note. NA = not available.

Measures

      Talk time. Throughout the group decision-making task, participants wore headsets attached to Zoom H1 Handy Portable Digital Recorders, which recorded their voices. The headsets we used have a built-in 90 degree X/Y microphone and support up to 24-bit/96kHz WAV audio. Talk time data was analyzed using a script written in Python, version 3.4. The script analyzed each participants' audio file by setting a threshold of intensity in decibels above which the participant's voice could be separated from the other group members' speech and from outside noise. The script then recorded the number of seconds that each participant spent talking within a 30-second time-frame and exported a separate file for each participant. A coder then listened to the first four segments (two minutes) and checked that the talk time exported by the script matched the amount of time that the participant spent talking. The script initially analyzed each ten milliseconds of the recording and checked whether the average intensity per ten ms was above the set threshold. However, after having the coder check the first four segments, we realized that the script was picking up too much noise in a ten ms unit, and we increased the unit to one second. Thus, the script analyzed each second and checked whether the average intensity, in decibels, was above the set threshold (threshold was, on average, 50 decibels). Thus, talk time for each participant is the number of seconds per 30-second interval that the script detected the average intensity to be above the set threshold.

      Video-coded behaviors. Throughout the group decision-making task, participants were video-recorded using handheld camcorders on tripods. One camera filmed each participant individually, and the videos of all group members were later merged together and synchronized. One research assistant, a master coder trained by the first author, coded videos of the group interactions for the following interactions. We had a second research assistant, who was trained

both by the first author and the master coder, code all of the videos as well in order to calculate

interrater reliabilities reported (see Table S2). We used the software Datavyu to code the videos.

We assessed interrater reliability of each behavior per 30-second interval of the group task using

two-way random effects single-measures, absolute agreement ICCs (Hallgren, 2012). After

assessing interrater reliability, we used the data from the master coder for analysis (we could not

average across both coders because the data are count data and need to be analyzed as such,

which would not be possible if the data from both coders were averaged).

Table S2

Intraclass correlations for coded behaviors.

	ICC
Interrupting someone	0.57
Getting interrupted by someone	0.56
Pitches in favor of a specific firm	0.73
Arguments against a specific firm	0.70

Interruptions. Coders watched the videos and indicated each time an interruption

occurred, as well as which participant did the interrupting and which participant was interrupted.

We did not consider an action to be an interruption if one participant was talking over another in

agreement with phrases such as "Yes", "I agree," or "You're right."

Pitches in favor of a firm. Coders marked the number of times that people advocated for

a specific search firm. For example, a participant might say "I think we should choose search

firm 2 because it has international offices which would help expand our search" or "I vote for search firm 4 because we have used them in the past."

     Arguments against a firm. Coders marked the number of times that people argued against a specific search firm. For example, a participant might say "I don't want to choose search firm 3 because they only have two senior recruiters" or "I vote against search firm 2 because it's too large."

     Desire to work together again. After the group decision-making task, we asked participants "How much would you want to work with Group Member [letter] on a similar task again?" on a scale of 1 (*not at all*) to 7 (*very much*).

Results

IBI Reactivity

     We estimated a linear growth curve model in which individual time points were nested within persons. We specified a random intercept, a random slope for time, and the within-person covariance between the two (i.e., the relationship between the random intercept and the random slope for time). We initially estimated a model that allowed for errors to be correlated between persons within the group within time points (i.e., a two-level crossed model) but this model would not converge. Thus, we included a third level (group) into the growth curve model (variance due to group was not significant, $p = .70$).

     In follow-up analyses, we examined whether IBI reactivity varied between successful and unsuccessful high-status women and found that it did not, $F(1, 26.00) = 0.07$, $p = .79$. We also examined whether IBI reactivity varied between low-status women and low-status men and found that it did not, $F(1, 84.17) = 0.01$, $p = .94$. Thus, IBI reactivity differed as a function of people's status assignment such that high-status women showed stronger reactivity than others,

and within status assignments, did not vary as a function of successful persuasion (in the high-status condition) or as a function of gender (in the low-status condition).

Perceptions of Persuasiveness

We specified a random effect for target and a random effect for perceiver. Additional analyses revealed that role did not interact with gender of the perceiver, $F(3, 266.15) = 1.47$, $p = .22$, nor was there a main effect of gender of the perceiver, $F(1, 176.25) = 0.01$, $p = .91$. Thus, high-status winners were perceived as more persuasive than all other types of group members by both male and female perceivers.

Physiological Linkage

Success is a group-level variable and linkage is a dyad-level variable. In multilevel modeling, outcomes cannot be at a higher level than predictors (in this case, the outcome cannot be at the group level with a predictor at the dyad level). To examine linkage scores without first averaging them at the level of the group (which would mean losing their original dyadic unit), we treat role (including success) as the predictor and linkage as the outcome.

Talk Time

Talk data were recorded in 30 second intervals, resulting in a nested structure of talk time nested within persons, and persons nested within group. Data were non-normally distributed, and so we analyzed data using GEE, specifying a Poisson distribution and an exchangeable correlation matrix. We did not find a significant main effect of role, $\chi^2(3) = 6.20$, $p = .10$, nor a significant main effect of time, $\chi^2(1) = 0.91$, $p = .34$. There was also no interaction between role and time, $\chi^2(3) = 0.27$, $p = .97$.

We also examined whether talk time (on average throughout the interaction) was associated with how much people showed linkage to others. We examined whether there was an

association between talk time and the extent to which people were "receivers" in models of physiological linkage and found that there was not, $\chi^2(1) = 0.16$, $p = .69$. We also examined whether there was an association between talk time and the extent to which people were "senders" in models of physiological linkage and found that there was not, $\chi^2(1) = 1.84$, $p = .18$.

Interruptions

Data were count data, and so we analyzed data using GEE, specifying a Poisson distribution and an exchangeable correlation matrix. In our first analysis, we analyzed whether there was an effect of role on the number of times that people interrupted others in the group. There was no main effect of role, $\chi^2(3) = 0.65$, $p = .89$. There was a main effect of time, $\chi^2(1) = 7.95$, $p = .005$, such that people interrupted more as time went on. There was no interaction between role and time, $\chi^2(3) = 0.83$, $p = .84$.

In a second analysis, we analyzed whether there was an effect of role on the number of times that people were interrupted by others in the group. There was no main effect of role, $\chi^2(3) = 1.53$, $p = .68$. There was a main effect of time, $\chi^2(1) = 11.13$, $p = .001$, such that people were interrupted more as time went on. There was no interaction between role and time, $\chi^2(3) = 3.51$, $p = .32$.

Pitches in Favor of a Firm

Data were count data, and so we analyzed data using GEE, specifying a Poisson distribution and an exchangeable correlation matrix. There was no main effect of role on the

number of pitches people made in favor of a firm, $\chi^2(3) = 3.49$, $p = .32$, nor a main effect of time, $\chi^2(1) = 0.36$, $p = .55$. There was no interaction between role and time, $\chi^2(3) = 5.03$, $p = .17$.

Arguments against a Firm

Data were count data, and so we analyzed data using GEE, specifying a Poisson distribution and an exchangeable correlation matrix. There was no main effect of role on the number of arguments people made against a firm, $\chi^2(3) = 3.43$, $p = .33$, nor was there a main effect of time, $\chi^2(1) = 1.06$, $p = .30$. There was only a marginally significant interaction between role and time, $\chi^2(3) = 6.70$, $p = .082$.

Desire to Work Together Again

We examined how much group members wanted to work with each other again, treating role as the predictor (4 levels: high-status successful women, high-status unsuccessful women, low-status women, and low-status men). A main effect of role was found, $F(3, 104.13) = 3.10$, $p = .03$. People wanted to work again with successful high-status women ($M = 5.31$, $SD = 1.26$) more than unsuccessful high-status women ($M = 4.35$, $SD = 1.40$), $t(109.42) = -2.89$, $p = .005$, low-status women ($M = 4.70$, $SD = 1.49$), $t(102.11) = -2.32$, $p = .022$, and low-status men ($M = 4.75$, $SD = 1.44$), $t(101.84) = -2.20$, $p = .030$. None of the other groups differed from one another (ps > .12). Additional analyses revealed that role did not interact with gender of the perceiver, $F(3, 285.35) = 1.07$, $p = .36$, nor was there a main effect of gender of the perceiver, $F(1, 164.97) = 0.03$, $p = .87$. Thus, both male and female group members wanted to work again with high-status winners more than all other types of group members.

Effect of Group Size

We conducted our primary analyses in the main text, while including a main effect of group size and interactions between all predictors in the main text and group size.

IBI reactivity. The main effect of group size was not significant, $F(1, 52.16) = 0.02$, $p = .89$. There was also no interaction between group size and status, $F(1, 93.24) = .001$, $p = .98$. There was no interaction between group size, status, and time, $F(1, 96.41) = 0.18$, $p = .67$.

Physiological linkage for receivers. The main effect of group size was not significant, $\chi^2(1) = .010$, $p = .92$. There was also no interaction between group size and condition on linkage, $\chi^2(3) = 2.72$, $p = .44$.

Physiological linkage for senders. The main effect of group size was marginally significant, $\chi^2(1) = 3.07$, $p = .08$, but there was no interaction between group size and condition on linkage, $\chi^2(3) = 4.28$, $p = .23$.

Self-reported persuasion. The main effect of group size was not significant, $F(1, 134.71) = 0.52$, $p = .47$. There was also no interaction between group size and condition, $F(3, 136.96) = 0.42$, $p = .74$.

Effect of Group Gender Composition

We conducted our primary analyses in the main text, while including a main effect of group gender composition and interactions between all predictors in the main text and group gender composition.

IBI reactivity. The main effect of gender composition of the group (ranged from 25% to 100%) was not significant, $F(1, 48.47) = 1.35$, $p = .25$. There was also no interaction between status (high vs. low) and gender composition of the group, $F(1, 93.06) = 0.01$, $p = .91$. There was no interaction between status, time, and gender composition, either, $F(1, 113.98) = 0.004$, $p = .95$.

Physiological linkage for receivers. The main effect of gender composition of the group was not significant, $\chi^2(1) = .02$, $p = .89$. There was also no interaction between condition as a 4-

level variable (consistent with what is reported in the main text) and gender composition of the group, $\chi^2(3) = 4.53$, $p = .21$.

Physiological linkage for senders. The main effect of gender composition of the group was not significant, $\chi^2(1) = 0.24$, $p = .63$. There was also no interaction between status as a 4-level variable (consistent with what is reported in the main text) and gender composition of the group, $\chi^2(3) = 1.42$, $p = .70$.

Self-reported persuasion. The main effect of gender composition of the group was not significant, $F(1, 105.83) = 0.09$, $p = .77$. There was also no interaction between condition and gender composition of the group on how persuasive a target was rated to be, $F(3, 110.97) = 0.78$, $p = .51$.

Effect of Familiarity

We conducted our primary analyses in the main text, while including a main effect of familiarity and interactions between all predictors in the main text and familiarity.

Linkage as receiver. For linkage as receiver, there was no main effect of whether the participant knew a group member or not on how strongly that person showed linkage to that group member, $\chi^2(1) = 1.21$, $p = .27$. There was also no interaction between condition and whether you knew the group member on linkage to that group member, $\chi^2(3) = 0.26$, $p = .97$.

Linkage as sender. For linkage as sender, there was a marginally significant main effect on whether the participant knew a group member on how much that they influenced that group member, $\chi^2(1) = 3.49$, $p = .06$; if a group member knew their partner, that partner showed stronger linkage to them than if they did not know their partner. There was no interaction

between condition and whether you knew the group member on how much that group member was linked to you, $\chi^2(3) = 1.44$, $p = .67$.

Self-reported persuasion. For ratings of persuasiveness, there was no main effect of knowing a group member on how persuasive that group member was perceived to be, $F(1, 314.13) = 1.60$, $p = .21$, nor did knowing a group member interact with condition to predict ratings of persuasiveness, $F(3, 314.74) = 0.12$, $p = .95$.

Effect of Gender Inequality Index

We conducted our primary analyses in the main text, while adjusting for receiver and sender GII.

IBI reactivity. The main effect of receiver GII was not significant, $F(1, 105.65) = 2.52$, $p = .12$. The main effect of sender GII was not significant, $F(1, 59.20) = 0.85$, $p = .36$. The main effect of status (reported in the main text) was still significant, $F(1, 88.08) = 4.72$, $p = .033$.

Physiological linkage for receivers. The main effect of receiver GII was not significant, $\chi^2(1) = 0.007$, $p = .93$. The main effect of sender GII was not significant, $\chi^2(1) = 0.04$, $p = .85$. The main effect of condition (reported in the main text) was still marginally significant, $\chi^2(3) = 6.46$, $p = .091$.

Physiological linkage for senders. The main effect of receiver GII was not significant, $\chi^2(1) = 0.28$, $p = .60$. The main effect of sender GII was not significant, $\chi^2(1) = 0.009$, $p = .92$.

Self-reported persuasion. The main effect of receiver GII was not significant, $F(1, 101.95) = 0.80$, $p = .38$. The main effect of sender GII was significant, $F(1, 102.79) = 9.61$, $p = .002$. People from countries with greater gender inequality rated their partners as more persuasive; however, the main effect of condition reported in the main text still remained significant, $F(3, 109.99) = 2.72$, $p = .048$.