

Online Supplemental Material for

**Listen, Follow Me:
Dynamic Vocal Signals of Dominance Predict
Emergent Social Rank in Humans**

ONLINE SUPPLEMENTAL MATERIAL FOR DYNAMIC VOCAL SIGNALS OF DOMINANCE PREDICT EMERGENT SOCIAL RANK IN HUMANS

TABLE OF CONTENTS

SUPPLEMENTAL EXPERIMENTAL PROCEDURES.....	4
Study 1.....	4
Assessing Behavioral Decision-Making Impact.....	4
Study 2.....	5
Experimental Stimuli and Participant Instructions.....	5
Table S1.....	6
SUPPLEMENTARY RESULTS AND ANALYSES.....	7
Study 1.....	7
Sample Demographic Information and Summary Statistics.....	7
Table S2.....	9
Missing Data.....	9
Alternative Model Specifications with Additional Controls.....	10
Table S3.....	12
Model Fit Information.....	13
Table S4.....	14
Do Men and Women Show Different Rank-Related Pitch Trajectories?.....	15
Table S5.....	18
Robustness Check using the Three Separate Indices of Social Rank.....	15
Table S6.....	19
Table S7.....	21
Assessing Change in Pitch with Difference Scores.....	22
Table S8.....	23
Figure S1.....	25
Absolute and Habitual Pitch and Social Rank.....	25
Pitch Modulations are Related to Dominance-Based Rank, But Not Prestige-Based Rank.....	26
Table S9.....	27

Table S10	30
Study 2.....	30
Forced Choice Summary Data.....	30
Figure S2	31
Forced Choice Results with Controls	31
Table S11	34
Table S12	34
SUPPLEMENTAL REFERENCES	35

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

These data derive from a larger project that was designed to examine the psychological underpinnings of social hierarchy. The current research complements this prior work, which focused on the behavioral strategies that promote rank (Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013) and the patterns of social attention in hierarchies (Foulsham, Cheng, Tracy, Henrich, & Kingstone, 2010), by addressing the novel question of how vocal changes influence the outcomes of rank contests.

Below we describe the calculation of our behavioral measure of social rank.

STUDY 1

ASSESSING BEHAVIORAL DECISION-MAKING IMPACT

Decision-making impact was measured by calculating the absolute difference between each participant's private ranking of each item on the Lost on the Moon task and his/her group's final ranking of that item, then summing across all 15 items and multiplying by -1, such that scores with a higher value (i.e., negative values closer to 0) reflect greater decisional impact. This scoring procedure can be represented as:

$$y_{ij} = -1(\sum_{k=1}^{15} |x_{ijk} - x_{jk}|)$$

where y_{ij} is the influence score of subject i from group j . x_{ijk} is subject i 's rating on item k . x_{jk} is group j 's rating on item k .

STUDY 2

EXPERIMENTAL STIMULI AND PARTICIPANT INSTRUCTIONS

In the master recording, a male research assistant read three statements into an Audio-Technica AT2020 microphone, positioned approximately 10cm away. Recordings were encoded in mono using Cool Edit Pro 2.1 at 44.1 kHz sampling rate and 16-bit quantization, and saved as three separate “wav” files. The statements are as follows:

Statement 1: “Hi, I’m Joshua. I’ve worked here at PIQ as a sales and marketing director for six and a half years. I’m glad to be part of this team that will get to decide how the company’s year-end bonus is to be allocated among our top 6 employees.”

Statement 2: “It’s not that this second candidate, Luke Myers, isn’t great. But he’s only been with the company for four months. I think people who’ve worked here longer and made more contributions deserve a larger bonus.”

Statement 3: “Ok. I agree with what Jon said. We should give a larger share of the bonus to this fifth candidate, Michael Lawson. Look at his performance over the last year. He outshines everyone else in how well he’s done.”

As outlined in the main text, these segments were then raised and lowered in pitch, independent of other acoustic properties to create one version with a progressively deepening pitch trajectory across statements 1 to 3, and another version that contains a progressively rising pitch trajectory. All pitch manipulations were carried out in Praat phonetic analysis software using the pitch synchronous overlap add algorithm, and verified using the autocorrelation algorithm. The mean pitch of these segments in the two recordings is summarized in Table S1. Note that the two stimuli are identical in all respects other than pitch, including the average pitch across the entire duration of each recording.

TABLE S1. MEAN SUMMARY PITCH STATISTICS OF VOICE STIMULI

	Segment 1 (Hz)	Segment 2 (Hz)	Segment 3 (Hz)	Mean (Hz)
Deepening pitch recording	124.17	105.15	95.63	108.32
Rising pitch recording	94.21	105.15	124.41	107.92

Both pairs of vocal stimuli were presented to participants, who received the following written instructions at the outset of the experiment:

“In this short survey, your task is to assist an employee who is trying to prepare for and rehearse a speech that he/she plans to give at the upcoming company meeting. In the meeting, this employee will be getting together with colleagues who, like him/her, are part of the company's year-end bonus allocation committee. Everyone present is the head of his/her specific division, and they will each be presenting their case for why the \$50,000 year-end bonus should be awarded to one specific top-performing employee in their division who is working directly under their supervision. The final decision will be based on everyone's collective consensus, so they are each motivated to influence others into endorsing their division's candidate.

Your task is to help one particular division head who, like the others, is trying to figure out how he/she should deliver his/her speech to most effectively influence the others. He/she has recorded two versions of how he/she could say what he/she wants to say to the committee. These two deliveries have identical content but differ in their speech styles.

You are asked to listen carefully to both versions of this division head's speech rehearsal, and then indicate your perceptions of what the two versions convey.

In both versions, he/she will give you a sample of the planned content at three different parts of his/her speech. The first segment features statements that he/she will make at the very beginning of the team's meeting, the second segment will be spoken at approximately 3 minutes into the meeting, and the third segment will be spoken at approximately 7 minutes into the meeting.

On the next pages, you will listen to both versions, and each will be played in the form of a single video. Please play both videos, listen and watch very carefully, and then respond to a number of questions.”

After listening to both pairs of recordings, participants were asked to respond to questions that would probe their perceptions of rank-seeking intentions and

formidability. For each, they were asked to choose which of the pair of recordings they considered more characteristic. Specifically, the following questions were presented in randomized order:

Rank-Seeking Intentions:

“In which version did he appear to be trying to gain power?”

“In which version did he appear to be trying to be the leader of the committee?”

“In which version did he appear to be trying to control the committee's decision-making?”

“In which version did he appear to be trying to be assertive?”

Dominance:

“In which version did he appear to try to intimidate the other committee members?”

“In which version did he appear to try to be threatening?”

Prestige:

“In which version did he appear to try to gain the admiration of other committee members?”

“In which version did he appear to try to gain the respect of other committee members?”

SUPPLEMENTARY RESULTS AND ANALYSES

Here, we report additional results and analyses to supplement the primary findings presented in the main text.

STUDY 1

SAMPLE DEMOGRAPHIC INFORMATION AND SUMMARY STATISTICS

Below, we first present mean demographic information and summary statistics on the main variables. Second, we examine the robustness of our primary results to the inclusion of a number of additional controls by building on the baseline model presented in the main text. These controls include age, height, weight, group size, and postural expansiveness. Third, we explore the robustness of our conclusions from the growth model

by using goodness of fit measures. Fourth, we present results of growth models using the three separate indices of rank—group member-rated rank, outside observer-rated rank, and behavioral decision-making impact—to supplement the results shown in the main text using the composite of these measures. Fifth, in another check of our conclusions, we present analyses that use a simple difference score approach to assess change in pitch. For these analyses, the pitch of each individual's first utterance was subtracted from the pitch of their third utterance, and these difference scores were regressed on social rank and controls. Finally, we examine the predictive capacity of absolute pitch levels (i.e., rather than change in pitch) on social rank outcomes. These analyses confirm that the observed differences between individuals in social rank cannot be reliably explained by absolute differences in pitch, and are instead a function of *changes* in vocal pitch.

TABLE S2. MEAN DESCRIPTIVE INFORMATION ON SAMPLE, PITCH PARAMETERS, AND SOCIAL RANK

	% of Sample	Mean	<i>SD</i>	Minimum	Maximum
Gender					
Men	54.01				
Women	45.99				
Ethnicity					
East Asian	46.52				
Caucasian	29.95				
South Asian	10.70				
Hispanic/Latino	1.60				
Mixed/Other	11.23				
Age		23.01	6.08	17	52
Group Size					
4-member group	14.97				
5-member group	26.74				
6-member group	54.55				
7-member group	3.74				
Pitch of 1 st utterance (Hz)					
Men		114.08	17.15	79.46	171.43
Women		201.30	32.54	109.10	277.05
Pitch of 2 nd utterance (Hz)					
Men		115.21	18.07	78.16	162.79
Women		206.61	28.10	161.11	274.29
Pitch of 3 rd utterance (Hz)					
Men		114.91	15.19	81.10	163.44
Women		206.98	26.49	155.78	285.70
Social rank composite		-.003	.78	-2.28	2.00
Group member-rated rank		4.12	1.12	1.59	6.76
Outside observer-rated rank		2.80	1.00	1	5
Behavioral decision-making impact		-38.03	13.29	-84	-8
Perceived dominance		2.33	.81	1.19	5.43
Perceived prestige		4.93	.62	3.09	6.36

MISSING DATA

Although a total of 191 participants completed the study, the video-recording of one session of 4 participants (and as a result their speech samples) was unavailable due to equipment failure, resulting in a total of 187 participants for whom audio

samples were available. However, vocal pitch parameters on all three utterances were indeterminable for 14 participants, who either did not speak in the group task or spoke too softly for speech to be recognized by the Praat phonetic analysis software. The analyses presented in the main text and below therefore contain data from 173 participants for whom any pitch parameters were available.

ALTERNATIVE MODEL SPECIFICATIONS WITH ADDITIONAL CONTROLS

In a series of individual growth models, we estimated the pitch of each utterance from time, social rank, and the interaction of time \times social rank, as well as additional controls added sequentially, including gender, age, height and weight, group size, and postural expansiveness. Gender was dummy coded, with women = 1. Self-reported height was measured in feet, and weight in pounds. Group size was captured by two dummy variables, group size 5 and group size 6 or 7, for comparison to the reference group size of 4. Postural expansiveness was obtained from detailed coding of each participant's postural displays exhibited during the group task from video-recordings. To minimize capturing moment-to-moment random fluctuations in participants' displays and instead derive a more precise measure of expansiveness, one research assistant, who was blind to hypotheses, selected eight 20-second segments from throughout each group's entire interaction. These segments were chosen to feature moments of key decisions and discussion in the group. Two other trained research assistants viewed these eight segments and independently coded the extent to which each participant "occupied room with the body", "held a wide posture", and "extended arms out from the body", on a 6-point Likert scale (0 = *Not at all present*; 5 = *Extreme intensity*). For each item, we summed the rating across all eight segments and then averaged across the two coders (i.e., aggregating across

the 8 ratings \times 2 coders = 16 observations for each participant). Finally, each individual's score on the three items were averaged to form an overall measure of postural expansiveness ($\alpha = .91$). All continuous predictors—age, height, weight, and postural expansiveness—were grand mean centered in the model.

To examine the effects of these controls, compare Models 1 (baseline, no controls) and 2 (with gender as control, reported in the main text) with Models 3-6 (with additional controls) in Table S3 below. The coefficients on the time \times social rank interaction term in Models 2-5, ranging from -4.80 to -4.32, remain stable from model to model and comparable to that of Model 2 (estimated to be -4.45), as well as statistically significant. In all models, apart from a significant gender main effect, none of the controls were significant at conventional levels. Model 2 showed the lowest BIC, indicating best fit adjusted for both model and sample size, so we reported results from this specification in the main text. Overall, these results converge to indicate that the diverging pitch trajectories found for low- and high-ranking individuals are unlikely to be driven by differences in age, gender, body size, or by dynamics of the local group, such as the number of group members present or one's physical expansiveness during the interaction. Taken together, results from all six models converge to indicate that, independent of these demographic and contextual factors, each 1-point increase in social rank was associated with a pitch trajectory that lowered approximately 4-5 Hz over each utterance.

TABLE S3. HIERARCHICAL LINEAR MODELS PREDICTING VOCAL PITCH FROM THE MAIN AND INTERACTIVE EFFECTS OF TIME AND SOCIAL RANK COMPOSITE, AS WELL AS CONTROL VARIABLES: AGE, GENDER (1 = WOMEN), HEIGHT, WEIGHT, GROUP SIZE DUMMIES, AND POSTURAL EXPANSIVENESS, AND THEIR INTERACTION WITH TIME. COEFFICIENTS ARE FOLLOWED BY STANDARD ERRORS IN PARENTHESES.

Predictor variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Time	1.82† (0.96)	0.86 (1.26)	0.89 (1.26)	1.45 (1.53)	3.60 (2.66)	4.12 (2.71)
Social Rank	9.93* (5.02)	5.85* (2.43)	5.84* (2.42)	7.00** (2.49)	6.93** (2.47)	8.88*** (2.59)
Time × Social Rank	-4.44*** (1.34)	-4.45*** (1.33)	-4.45*** (1.32)	-4.39** (1.37)	-4.32** (1.37)	-4.80*** (1.43)
Gender (Female=1)		86.78*** (3.61)	86.96*** (3.62)	80.45*** (5.05)	80.70*** (5.08)	82.54*** (4.99)
Time × Gender		2.48 (1.89)	2.42 (1.89)	1.14 (2.67)	1.28 (2.71)	0.66 (2.70)
Age			-0.19 (0.30)	-0.04 (0.31)	-0.12 (0.32)	-0.34 (0.32)
Time × Age			0.07 (0.15)	0.06 (0.16)	0.09 (0.17)	0.16 (0.17)
Height (feet)				-8.96 (7.59)	-8.75 (7.52)	-7.84 (7.35)
Weight (lbs)				-0.06 (0.07)	-0.04 (0.07)	-0.02 (0.07)
Time × Height				-4.40 (3.98)	-4.43 (3.97)	-4.80 (3.95)
Time × Weight				0.02 (0.03)	0.01 (0.04)	0.01 (0.04)
Group Size 5 (dummy)					8.79 (5.98)	8.44 (6.04)
Group Size 6 or 7 (dummy)					9.38† (5.05)	8.61† (5.09)
Time × Group Size 5					-3.27 (3.13)	-3.47 (3.22)
Time × Group Size 6 or 7					-2.36 (2.65)	-2.35 (2.71)

Postural Expansiveness						-0.26 (0.46)
Time × Postural Expansiveness						-0.01 (0.25)
(Intercept)	152.31*** (3.75)	114.08*** (2.41)	113.99*** (2.41)	117.00*** (2.89)	109.14*** (5.08)	108.24*** (5.09)
AIC	4678.64	4367.45	4371.04	4369.75	4373.87	4359.55
BIC	4712.16	4409.35	4421.32	4436.79	4457.68	4451.69
Observations	488	488	488	488	488	488
Clusters	173	173	173	173	173	172

† p<0.10, * p<0.05, ** p<.01, *** p<.001

MODEL FIT INFORMATION

To further support the analyses reported in Table S3 and the main text, which tested the significance of the time × social rank cross-level interaction using null hypothesis testing (i.e., testing whether the ratio of the coefficient to the standard error differs from zero; Raudenbush & Bryk, 2002), we examine model fit here. We compared the Akaike information criterion (AIC) and Bayesian information criterion (BIC) of models that include the cross-level time × social rank interaction for each of the five specifications, presented above in Table S3, with their corresponding nested models in which the interaction term is omitted.

To examine the effects of the time × social rank interaction, compare the AIC and BIC of Models 1 to 6 reported in Table S3 above with the fit indices of Models 1' to 6' in Table S4 below. Across all six specifications, the AIC and BIC are lower in the original model in Table S3 above with the interaction term, than in the nested model in Table S4 without the

interaction term, indicating better fit of the models with the interaction term, across the board. The differences in AIC range from 9.12 to 8.17 and the differences in BIC range from 4.94 to 3.62, all of which exceed the threshold value, 2, for positive evidence indicating a difference in fit (Raftery, 1995). These results reconfirm the predictive importance of the time \times social rank interaction, and by implication the conclusion that higher social rank is associated with a deepening pitch profile over time.

TABLE S4. HIERARCHICAL LINEAR MODELS PREDICTING VOCAL PITCH FROM THE MAIN EFFECTS OF TIME AND SOCIAL RANK COMPOSITE, AS WELL AS CONTROL VARIABLES INCLUDING AGE, GENDER, HEIGHT, WEIGHT, GROUP SIZE DUMMIES, AND POSTURAL EXPANSIVENESS, AND THEIR INTERACTION WITH TIME. COEFFICIENTS ARE FOLLOWED BY STANDARD ERRORS IN PARENTHESES. MODELS 1' TO 6' HERE, WHICH DO NOT INCLUDE THE TIME \times SOCIAL RANK INTERACTION TERM, CORRESPOND TO MODELS 1 TO 6 IN TABLE S3 ABOVE THAT CONTAIN THE INTERACTION. FOR BOTH FIT INDICES (AIC AND BIC), LOWER VALUES INDICATE BETTER FIT.

	Model 1'	Model 2'	Model 3'	Model 4'	Model 5'	Model 6'
Time	1.05 (0.95)	0.38 (1.28)	0.41 (1.28)	1.64 (1.55)	3.88 (2.70)	4.93† (2.74)
Social Rank	7.02 (4.94)	1.13 (1.98)	1.13 (1.98)	2.21 (2.00)	2.28 (1.99)	3.67† (2.09)
Gender (Female=1)		87.29*** (3.64)	87.48*** (3.65)	82.59*** (5.04)	82.64*** (5.08)	84.16*** (5.01)
Time \times Gender		2.05 (1.92)	1.99 (1.92)	-0.62 (2.66)	-0.33 (2.71)	-0.94 (2.70)
Age			-0.20 (0.30)	-0.10 (0.32)	-0.18 (0.32)	-0.38 (0.32)
Time \times Age			0.06 (0.16)	0.10 (0.17)	0.13 (0.17)	0.20 (0.17)
Height (feet)				-7.43 (7.64)	-7.31 (7.56)	-6.56 (7.41)
Weight (lbs)				-0.04 (0.07)	-0.01 (0.07)	0.01 (0.07)
Time \times Height				-5.69 (4.03)	-5.66 (4.02)	-6.12 (3.99)
Time \times Weight				0.00 (0.03)	-0.01 (0.04)	-0.02 (0.04)
Group Size 5 (dummy)					9.48	10.01†

					(6.02)	(6.07)
Group Size 6 or 7 (dummy)					9.48†	9.49†
					(5.09)	(5.13)
Time × Group Size 5					-3.81	-4.80
					(3.18)	(3.25)
Time × Group Size 6 or 7					-2.37	-3.01
					(2.69)	(2.75)
Postural Expansiveness						0.03
						(0.45)
Time × Postural Expansiveness						-0.25
						(0.24)
(Intercept)	152.59***	114.31***	114.21***	116.45***	108.43***	107.08***
	(3.76)	(2.43)	(2.43)	(2.91)	(5.11)	(5.12)
<i>AIC</i>	4687.42	4376.52	4380.09	4377.92	4381.68	4368.67
<i>BIC</i>	4716.75	4414.23	4426.19	4440.77	4461.30	4456.63
Observations	488	488	488	488	488	487
Clusters	173	173	173	173	173	172

† $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

DO MEN AND WOMEN SHOW DIFFERENT RANK-RELATED PITCH TRAJECTORIES?

Do men and women use different kinds of vocal pitch-based strategies to signal and compete for social rank? Cultural explanations of gender differences, such as the notion of “gender code”, predict a sex difference in this respect (Cartei, Cowles, & Reby, 2012); specifically, cultural beliefs, practices, and institutions expect and reward the demonstration of masculinity in men and femininity in women, and also link masculinity (and the display of masculine traits and characteristics) with success and competition. These strong cultural associations may result in men and women conforming to the gender code by systematically manipulating their vocal pitch to be consistent with gender norms.

This account predicts that whereas men may readily acquire—by trial and error, imitation, or learning—the propensity to flaunt signs of masculinity, including deepening their voice in social situations, women may raise their pitch or show little systematic vocal modulation.

To test this account, we examined whether the association between pitch trajectories and social rank is moderated by the signaler's gender. It is important to note, however, that these analyses should be interpreted with caution, due to the limited statistical power afforded by our relatively small sample of men ($n = 96$) and women ($n = 77$) for precisely estimating within-gender effects in individual growth models. Statistical power and sample size concerns in estimating cross-level interactions in multilevel models are particularly complex and jointly influenced by interactive relationships among multiple factors, many of which are not relevant in single-level analyses (Mathieu, Aguinis, Culpepper, & Chen, 2012; Scherbaum & Ferreter, 2009; Snijders, 2005; Snijders & Bosker, 1993, 1994). Simulation studies (Bassiri, 1988; van der Leeden & Busing, 1994) suggest that whereas cross-level interactions generally require 30 observations at both the higher and lower level, fewer observations at either level leads to a rapid decline of power and precision for detecting cross-level interactions, such that with only five observations at the lower level, 150 upper level units are needed (for a review, see Kreft & de Leeuw, 1998). This suggests that, for our growth model, which has 3 repeated observations of pitch per individual, at least 150 subjects are needed in order to detect and estimate the cross-level time \times social rank interaction—a criterion that is met by our combined sample size of 173 participants, but not by either sample of men or women alone. Thus, any estimated effects of within-gender relationships may also be biased (Snijders & Bosker, 1993).

Nevertheless, results from a growth model that accounts for the possibility of gender moderation indicate no qualitative difference between men and women in their pitch strategies. Although the growth model revealed a significant time \times social rank \times gender interaction in predicting vocal pitch ($b = -4.67$, $SE = 2.19$, $p = .033$, $.95CI[-8.97, -.37]$; see Table S5 below for full model results), descriptively the pitch profiles for high- and low-ranking individuals are similar across both genders. Specifically, the model estimates a 3.40 Hz *increase* per utterance for low-ranking men, and a 11.97 Hz increase per utterance for low-ranking women, with a significantly steeper trajectory for women than for men ($p = .012$). Moreover, the model estimates a 1.74 Hz *decrease* per utterance for high-ranking men and a 2.61 Hz *decrease* for high-ranking women, and these two trajectories do not differ significantly ($p = .722$). Thus, these analyses indicate that for both men and women, a lowering pitch profile is associated with greater social rank, whereas a raising pitch profile is associated with lower social rank. Furthermore, the only gender difference that did emerge suggests that low-ranking women raise their pitch to a greater degree than their low-ranking male counterparts, whereas high-ranking men and women lower their pitch to a similar degree.

In sum, although these results are in need of future replication with studies using larger samples, these analyses offer very preliminary support for the notion that pitch change signals operate similarly across gender. These results are consistent with prior work indicating that lower voices function to increase dominance in women (as well as men), including findings that lower pitched female voices are perceived as more dominant and successful in acquiring resources (e.g., Apicella & Feinberg, 2009; Borkowska & Pawlowski, 2011; Jones, Feinberg, DeBruine, Little, & Vukovic, 2010), and that speaking in

a lower pitch (following experimental instructions) enhances women's subjective reports of power and dominance, as it does in men (Stel, Dijk, Smith, Dijk, & Djalal, 2012). The finding that low-ranking women raise their voice to a greater magnitude than similarly ranked men may be due to anatomical differences and constraints that equip women with a wider possible range of pitch (e.g., Cartei et al., 2012).

TABLE S5. HIERARCHICAL LINEAR MODELS PREDICTING VOCAL PITCH FROM THE MAIN EFFECTS OF TIME, SOCIAL RANK COMPOSITE, GENDER, THE INTERACTION AMONG THESE VARIABLES. COEFFICIENTS ARE FOLLOWED BY STANDARD ERRORS IN PARENTHESES. THIS MODEL CORRESPONDS TO MODEL 2 IN TABLE S3 ABOVE BUT WITH THE ADDITION OF THE TIME \times SOCIAL RANK \times GENDER INTERACTION TERM.

Predictor variables	Regression Coefficient (SE)
Time	0.61 (1.26)
Social Rank	5.86* (2.43)
Gender (female = 1)	86.73*** (3.62)
Time \times Social Rank	-2.57 (1.59)
Time \times Gender	3.39 (1.92)
Time \times Social Rank \times Gender	-4.67* (2.19)
(Intercept)	114.10*** (2.41)
Observations	488
Clusters	173

* $p < 0.05$, ** $p < .01$, *** $p < .001$

ROBUSTNESS CHECK USING THE THREE SEPARATE INDICES OF SOCIAL RANK

To check the robustness of our findings to each of the three specific indices that, in the main text, were aggregated into a social rank composite, and paralleling the results

reported above, we examined two specifications, one of which controls for gender (the best fitting specification shown in the main text and Model 2 in Table S3) and the other of which includes the full set of controls (Model 6 in Table S3).

Tables S6 and S7 display these two sets of individual growth models. Coefficients on the time \times rank index term in the models using outside observer-rated rank and behavioral decision-making impact were sizable, negative, and significant, but were not well estimated for group member-rated rank ($ps = .051$ and $.074$, respectively). Nevertheless, these results are generally consistent with each other and with findings based on the rank composite, indicating that our conclusions are robust across diverse approaches to measuring social rank. Note that the magnitude of the coefficients on social rank cannot be directly compared because units vary depending on the specific rank index.

TABLE S6. HIERARCHICAL LINEAR MODELS PREDICTING VOCAL PITCH FROM THE MAIN AND INTERACTIVE EFFECTS OF TIME AND SOCIAL RANK INDEX (WHICH DIFFERS IN EACH MODEL), AS WELL AS THE CONTROLS GENDER AND ITS INTERACTION WITH TIME. COEFFICIENTS ARE FOLLOWED BY STANDARD ERRORS IN PARENTHESES.

	Social Rank Index		
	Group member-rated rank	Outside observer-rated rank	Behavioral decision-making impact
Time	0.67 (1.28)	1.11 (1.28)	0.28 (1.27)
Social Rank	2.42 (1.71)	3.77† (1.94)	0.30* (0.14)
Time \times Social Rank	-1.79† (0.92)	-3.30** (1.07)	-0.19** (0.07)
Gender (Female=1)	87.06*** (3.64)	87.41*** (3.62)	86.64*** (3.65)
Time \times Gender	2.26 (1.91)	2.03 (1.89)	2.59 (1.91)
(Intercept)	114.18***	113.85***	114.56***

	(2.43)	(2.43)	(2.43)
Observations	488	488	485
Clusters	173	173	172

† $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE S7. HIERARCHICAL LINEAR MODELS PREDICTING VOCAL PITCH FROM THE MAIN AND INTERACTIVE EFFECTS OF TIME AND SOCIAL RANK INDEX (WHICH DIFFERS IN EACH MODEL), AS WELL CONTROL VARIABLES AGE, GENDER, HEIGHT, WEIGHT, GROUP SIZE DUMMIES, AND POSTURAL EXPANSIVENESS, AND THEIR INTERACTION WITH TIME. COEFFICIENTS ARE FOLLOWED BY STANDARD ERRORS IN PARENTHESES.

Predictor variables	Measure of Social Rank		
	Group member-rated rank	Outside observer-rated rank	Behavioral decision-making impact
Time	4.76† (2.74)	5.17† (2.72)	3.77 (2.77)
Social Rank	3.93* (1.80)	5.62** (2.06)	0.37** (0.14)
Time × Social Rank	-1.74† (0.97)	-3.53** (1.15)	-0.20** (0.07)
Age	-0.42 (0.32)	-0.36 (0.32)	-0.32 (0.32)
Gender (Female=1)	83.54*** (5.08)	84.18*** (5.00)	83.59*** (5.06)
Time × Age	0.20 (0.17)	0.17 (0.17)	0.15 (0.17)
Time × Gender	-0.13 (2.73)	-0.15 (2.68)	0.07 (2.72)
Height (feet)	-7.93 (7.52)	-6.74 (7.42)	-5.86 (7.46)
Weight (lbs)	0.01 (0.07)	-0.00 (0.07)	-0.00 (0.07)
Time × Height	-5.07 (4.02)	-5.32 (3.96)	-5.88 (3.98)
Time × Weight	-0.01 (0.04)	0.00 (0.04)	-0.00 (0.04)
Group Size 5 (dummy)	9.89 (6.12)	10.39† (6.06)	8.67 (6.15)
Group Size 6 or 7 (dummy)	9.82† (5.16)	10.01† (5.13)	7.96 (5.20)
Time × Group Size 5	-4.33 (3.25)	-4.37 (3.22)	-3.60 (3.26)
Time × Group Size 6 or 7	-2.97 (2.74)	-3.06 (2.72)	-2.03 (2.76)
Postural Expansiveness	-0.04 (0.46)	-0.21 (0.47)	0.11 (0.45)
Time × Postural	-0.15	0.00	-0.20

Expansiveness	(0.25)	(0.25)	(0.24)
(Intercept)	107.00*** (5.16)	106.29*** (5.12)	108.71*** (5.23)
Observations	487	487	484
Clusters	172	172	171

† $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

ASSESSING CHANGE IN PITCH WITH DIFFERENCE SCORES

As an alternative check of our results, we conducted analyses using the simple difference between the pitch of the third utterance and first utterance—the two available points that differed maximally in time—to assess change (Rogosa & Willett, 1983). We calculated a measure of change in pitch for each individual by subtracting the pitch of the first utterance from the pitch of the third utterance. If a group member raised her pitch over time, her pitch change score would take on a positive value, whereas if she lowered her pitch, her change score would take on a negative value.

Table S8 shows a series of linear regression models regressing pitch change on the social rank composite (grand mean centered), and sequentially added controls. Independent of the controls, none of which were significant, the coefficients on social rank were negative and significant at conventional levels across all seven models. A 1-point increase in social rank was associated with a decrease in pitch ranging from roughly -8.1 to -8.7 Hz. This corresponds to approximately a 6% change in fundamental frequency on average for each 1-point increase in social rank, a magnitude that prior perceptual studies have revealed to meet the discrimination threshold for humans, which for this frequency range generally ranges from 2 to 7% conditional on the

specific type of voice material (see Ives, Smith, & Patterson, 2005; Pisanski & Rendall, 2011; Puts, Hodges, Cárdenas, & Gaulin, 2007; Sinnott, Owren, & Petersen, 1987; Smith, Patterson, Turner, Kawahara, & Irino, 2005), and is thus perceptually relevant. Note, however, that this measured magnitude of change reflects pitch modulations observed over the initial vocalizations of the interaction. Given the brief time frame over which change is examined, our measure likely underestimates the true strength of the pitch change signals deployed. The actual salience and strength of pitch changes in rank contests, and whether these signals can be perceptually discriminated, is an important question for future work (see the General Discussion in the main text for further discussion). Nevertheless, these models estimate that, independent of the controls, the average low-ranking individual who scores at the 10th percentile on social rank raises his/her pitch by approximately 12.13 to 12.82 Hz from the first to third utterance. In contrast, the average high-ranking individual who scores at the 90th percentile on social rank *lowers* his/her pitch by approximately 4.53 to 4.96 Hz from the first to third utterance.

TABLE S8. ORDINARY LEAST SQUARE REGRESSIONS FOR SIMPLE DIFFERENCE BETWEEN PITCH AT THE THIRD AND FIRST UTTERANCES, WITH SOCIAL RANK. COEFFICIENTS ARE FOLLOWED BY STANDARD ERRORS IN PARENTHESES.

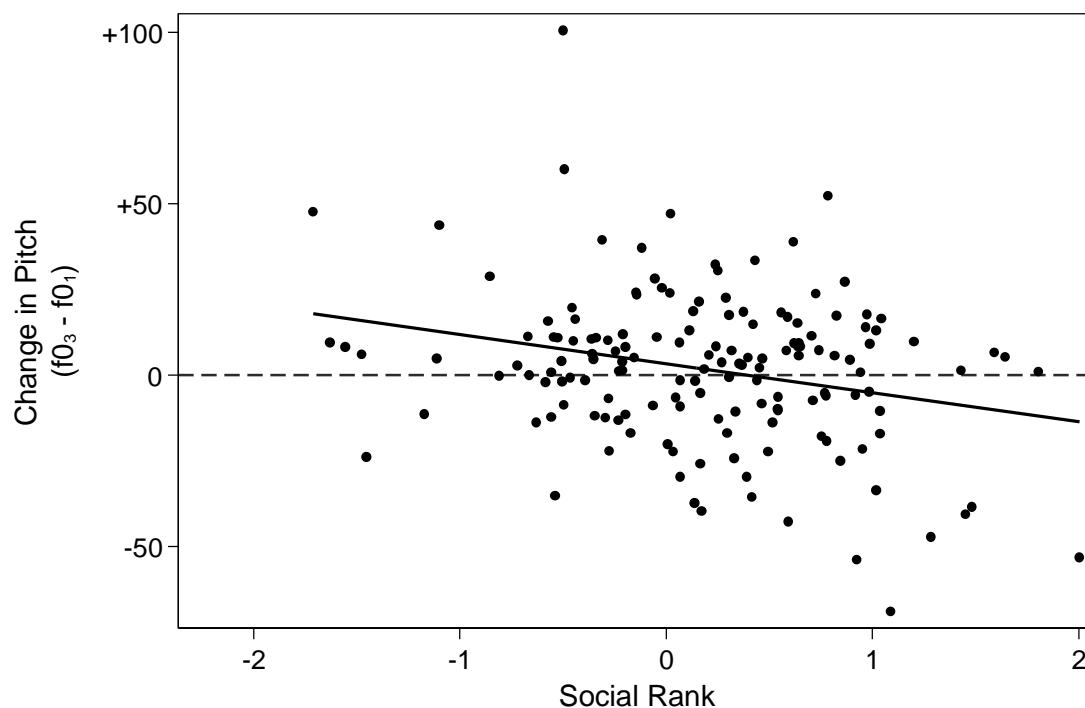
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Social Rank (Composite)	-8.48** (2.56)	-8.63*** (2.57)	-8.65*** (2.57)	-8.48** (2.65)	-8.31** (2.66)	-8.11** (2.78)	-8.11** (2.78)
Gender (Female=1)		2.80 (3.56)	2.58 (3.58)	-0.01 (5.09)	0.51 (5.17)	0.46 (5.19)	0.46 (5.19)
Age			0.22 (0.29)	0.20 (0.31)	0.26 (0.31)	0.26 (0.32)	0.26 (0.32)
Height (feet)				-8.50 (7.51)	-8.22 (7.52)	-8.32 (7.55)	-8.32 (7.55)

Weight (lbs)				0.04 (0.06)	0.02 (0.07)	0.02 (0.07)	0.02 (0.07)
Group Size 5 (dummy)					-8.18 (5.90)	-8.58 (6.13)	-8.58 (6.13)
Group Size 6 or 7 (dummy)					-5.63 (4.99)	-5.95 (5.16)	-5.95 (5.16)
Postural Expansiveness						-0.12 (0.47)	-0.12 (0.47)
(Intercept)	3.38† (1.83)	2.17 (2.39)	2.26 (2.40)	3.39 (2.90)	8.50† (5.01)	8.80† (5.16)	8.80† (5.16)
<i>AIC</i>	1389.27	1390.64	1392.06	1394.68	1396.60	1398.53	1398.53
<i>BIC</i>	1395.34	1399.75	1404.21	1412.90	1420.90	1425.86	1425.86
<i>N</i>	154	154	154	154	154	154	154

† $p < 0.10$, * $p < 0.05$, ** $p < .01$, *** $p < .001$

Figure S1 plots change in pitch—which, as can be seen, shows sizable variation between individuals—as a function of social rank. Confirming the results from the individual growth models reported in the main text and above, high-ranking individuals tended to deepen their pitch over the course of the interaction, whereas low ranking individuals raised their pitch.

FIGURE S1. SCATTER PLOT (WITH BEST-FITTING REGRESSION LINE) OF CHANGE IN VOCAL PITCH AS A FUNCTION OF SOCIAL RANK. CHANGE IN PITCH WAS COMPUTED BY SUBTRACTING THE PITCH OF AN INDIVIDUAL'S FIRST UTTERANCE FROM THE PITCH OF HIS OR HER THIRD UTTERANCE. THE DOTTED LINE SEPARATES INDIVIDUALS WHO LOWERED THEIR PITCH (BELOW THIS LINE) FROM THOSE WHO RAISED THEIR PITCH (ABOVE THIS LINE).



Note. On average, each 1-point increase in social rank was associated with a 8.48 Hz drop in vocal pitch. This negative association remains significant when the participant who showed the greatest pitch change (+100.42 Hz) is removed from the analysis; the model still estimates a 7.6 Hz drop in pitch for each unit of social rank ($p = .002$).

ABSOLUTE AND HABITUAL PITCH AND SOCIAL RANK

To examine whether mean pitch levels—as opposed to *changes* in pitch—are related to rank, we correlated the rank composite with the pitch parameter of each utterance and with mean pitch across all three utterances. With one exception, all correlations were non-significant at conventional levels: $f0_1$ ($r = .05$, $p = .630$ for males), $f0_2$ ($r = .10$, $p = .346$ for males; $r = .003$, $p = .981$ for females), $f0_3$ ($r = -.02$, $p = .881$ for

males; $r = -.17$, $p = .179$ for females), and $f0_{\text{mean}}$ (reflecting habitual pitch; $r = -.01$, $p = .951$ for males; $r = .18$, $p = .13$ for females). However, the first utterance for females was significant, though in the opposite direction than expected ($r = .28$, $p = .013$). These results, combined with findings above, indicate that little of the variation between individuals in rank is explained by higher or lower mean momentary pitch levels or habitual pitch. Instead, rank is more directly linked to dynamics shifts in pitch over time.

PITCH MODULATIONS ARE RELATED TO DOMINANCE-BASED RANK, BUT NOT PRESTIGE-BASED RANK

Our analyses indicated that, as expected, individuals whose speech was characterized by a progressively deepening pitch profile attained greater dominance-based rank. Also as expected, prestige-based rank was not reliably associated with pitch change. In a growth model, pitch was modeled as a function of time of utterance, peers' ratings of dominance, peers' ratings of prestige, and the cross-level interaction between time and each of these rank-attainment strategies. The full series of models with a variety of controls are displayed in Table S9.

In the specification with only gender as control (Model 2; which yielded the lowest model BIC among all specifications and was reported in the main text), the time \times dominance interaction was significant and indicates a 2.84 Hz decrease in pitch per utterance for each 1-point increase in dominance. In contrast, the time \times prestige interaction was not significant. In the other models with controls (Models 3-6) and without controls (Model 1), the coefficient on time \times dominance remained negative, varying from -2.80 to -3.04. As above, none of the time \times prestige coefficients were significant. Notably, the pattern of results is robust across all of these alternative specifications.

TABLE S9. HIERARCHICAL LINEAR MODELS PREDICTING VOCAL PITCH FROM THE MAIN AND INTERACTIVE EFFECTS OF TIME, DOMINANCE, AND PRESTIGE, AS WELL AS CONTROL VARIABLES AGE, GENDER, HEIGHT, WEIGHT, GROUP SIZE DUMMIES, AND POSTURAL EXPANSIVENESS, AND THEIR INTERACTION WITH TIME. COEFFICIENTS ARE FOLLOWED BY STANDARD ERRORS IN PARENTHESES.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Time	1.54 (0.96)	0.94 (1.29)	1.03 (1.29)	1.84 (1.55)	3.73 (2.70)	4.49 (2.75)
Dominance	-0.41 (4.61)	3.81† (2.23)	4.23† (2.27)	4.83* (2.28)	4.53* (2.27)	5.34* (2.33)
Prestige	10.12 (6.41)	1.59 (3.11)	1.19 (3.13)	2.13 (3.18)	2.93 (3.21)	4.19 (3.19)
Time × Dominance	-2.80* (1.16)	-2.84* (1.17)	-3.04* (1.20)	-2.86* (1.21)	-2.83* (1.22)	-2.93* (1.26)
Time × Prestige	-1.38 (1.64)	-1.07 (1.65)	-0.89 (1.66)	-0.56 (1.69)	-0.55 (1.73)	-0.93 (1.72)
Gender (Female=1)		87.74*** (3.67)	88.15*** (3.68)	82.51*** (5.10)	82.28*** (5.12)	84.10*** (5.07)
Time × Gender		1.69 (1.92)	1.50 (1.93)	-0.35 (2.69)	0.02 (2.72)	-0.57 (2.71)
Age			-0.30 (0.31)	-0.19 (0.32)	-0.26 (0.32)	-0.48 (0.33)
Time × Age			0.14 (0.16)	0.15 (0.17)	0.18 (0.17)	0.25 (0.17)
Height (feet)				-9.23 (7.81)	-9.42 (7.74)	-8.99 (7.60)
Weight (lbs)				-0.04 (0.07)	-0.01 (0.07)	0.02 (0.07)
Time × Height				-4.58 (4.09)	-4.54 (4.09)	-4.80 (4.07)
Time × Weight				0.01 (0.03)	-0.00 (0.04)	-0.01 (0.04)
Group Size 5 (dummy)					9.54 (6.03)	9.93 (6.10)
Group Size 6 or 7 (dummy)					9.33† (5.18)	9.20† (5.25)
Time × Group Size 5					-3.64 (3.16)	-4.17 (3.24)
Time × Group Size 6 or 7					-1.90 (2.73)	-2.21 (2.81)
Postural Expansiveness						-0.09 (0.46)
Time × Postural						-0.09

Expansiveness						
(Intercept)	152.58***	113.87***	113.67***	116.26***	108.41***	107.02***
	(3.78)	(2.44)	(2.44)	(2.91)	(5.13)	(5.16)
<i>AIC</i>	4686.68	4376.78	4379.73	4378.54	4382.19	4369.61
<i>BIC</i>	4728.59	4427.06	4438.40	4453.97	4474.38	4470.13
<i>N</i>	488	488	488	488	488	487

† $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Moreover, additional analyses indicate that the effect of a deepening pitch profile in predicting greater dominance-based rank did not depend upon obtaining a specific level of prestige. When a three-way interaction term (time \times dominance \times prestige) was added to the same baseline model as above, which used only gender as control (Model 2), the three-way interaction was not statistically significant ($b = 1.44$, $SE = 1.38$, $p = .295$, $.95CI[-1.25, 4.14]$). Moreover, replicating our previous analysis, the time \times dominance interaction term remained a significant negative predictor of vocal pitch ($b = -2.63$, $SE = 1.19$, $p = .027$, $.95CI[-4.96, -.29]$), whereas the time \times prestige term had a null effect ($b = -1.35$, $SE = 1.66$, $p = .418$, $.95CI[-4.61, 1.92]$). These results offer further support to the notion that the link between pitch modulations and emergent social rank is underpinned by increased perceptions of dominance, but not prestige.

As a robustness check, we next turned to the alternative operationalization of pitch change using simple difference scores. We estimated a series of linear regression models in which pitch change—calculated by subtracting pitch of the first utterance from the third utterance—was regressed on dominance and prestige (both grand mean centered), along with, in some models, additional controls. While these models in Table S10 below show no effects of prestige or any of the controls, the coefficient on dominance remained significant and negative across all specifications, except in Model 6 where the effect was marginally

significant ($p = .053$) when the full set of controls was included. The coefficients on dominance indicate that a 1-point increase in dominance predicts a decrease in vocal pitch between 4.79 Hz and 5.54 Hz, independent of the controls. These models estimate that, on average, as dominance increases from the 10th percentile to the 90th percentile, the estimated pitch trajectory changes from a rising profile, marked by an approximate increase between 6.49 and 7.22 Hz, to a lowering profile, in which pitch *lowers* by approximately -1.61 to -2.14 Hz.

TABLE S10. ORDINARY LEAST SQUARE REGRESSIONS PREDICTING PITCH CHANGE, OPERATIONALIZED AS PITCH ON THE THIRD UTTERANCE MINUS PITCH ON THE FIRST UTTERANCE, FROM DOMINANCE AND PRESTIGE. COEFFICIENTS ARE FOLLOWED BY STANDARD ERRORS IN PARENTHESES.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Dominance	-5.10*	-5.02*	-5.54*	-5.16*	-5.14*	-4.79†
	(2.24)	(2.26)	(2.30)	(2.34)	(2.35)	(2.45)
Prestige	-2.11	-2.18	-1.80	-1.08	-1.22	-1.13
	(3.15)	(3.17)	(3.18)	(3.25)	(3.33)	(3.35)
Gender (Female=1)		1.24	0.76	-2.97	-1.98	-1.98
		(3.67)	(3.69)	(5.16)	(5.23)	(5.24)
Age			0.34	0.37	0.44	0.43
			(0.31)	(0.32)	(0.33)	(0.33)
Height (feet)				-8.89	-8.40	-8.65
				(7.81)	(7.81)	(7.84)
Weight (lbs)				0.01	-0.01	-0.01
				(0.07)	(0.07)	(0.07)
Group Size 5 (dummy)					-9.05	-9.84
					(6.02)	(6.22)
Group Size 6 or 7 (dummy)					-4.87	-5.58
					(5.21)	(5.40)
Postural Expansiveness						-0.25
						(0.48)
(Intercept)	2.67	2.12	2.35	3.97	8.70†	9.31†
	(1.85)	(2.47)	(2.48)	(2.97)	(5.15)	(5.29)
<i>AIC</i>	1396.75	1398.63	1399.33	1401.93	1403.54	1405.25
<i>BIC</i>	1405.86	1410.78	1414.52	1423.19	1430.87	1435.62
<i>N</i>	154	154	154	154	154	154

† $p < 0.10$, * $p < 0.05$, ** $p < .01$, *** $p < .001$

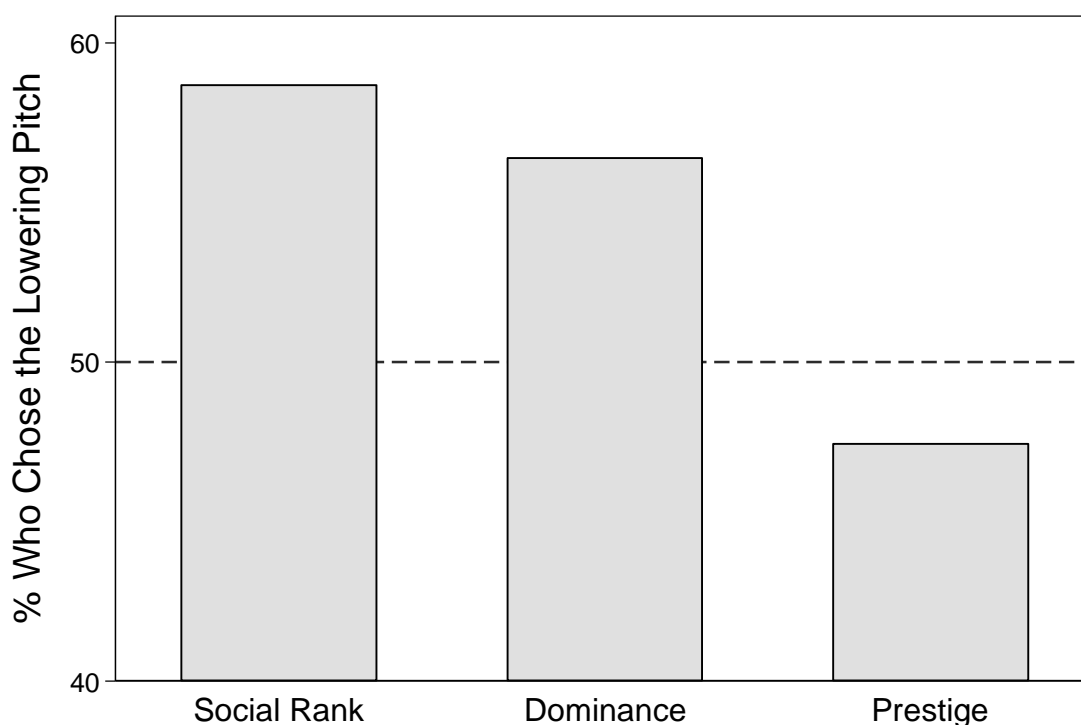
STUDY 2

FORCED CHOICE SUMMARY DATA

Figure S2 summarizes the data from Study 2. As can be seen, the proportion of participants who matched the social rank composite and the dominance composite to the

deepening pitch recording (compared to the alternative rising pitch recording) was significantly greater than would be expected by chance (i.e., 50%).

FIGURE S2. THE PROPORTION OF PARTICIPANTS WHO SELECTED THE DEEPENING PITCH RECORDING FOR EACH TRAIT. THE DOTTED LINE INDICATES CHANCE



FORCED CHOICE RESULTS WITH CONTROLS

To test whether these effects are robust to the inclusion of subject-level controls, we regressed the summative measure of each participant's average tendency to select the deepening voice on gender, age, method of recruitment (student or Mechanical Turk worker), and stimuli presentation order. With the exception of age, which was the only continuous predictor, all dichotomous predictors were represented by effects coding. Table S10 shows three regression models using these controls to predict the average tendency to select the deepening voice for communicating greater rank-

seeking intention, dominance, and prestige. With effects coding, the intercept in each model summarizes the model predicted unweighted grand mean preference for a deepening voice (coded '0' for each individual item) across all participants, accounting for the effects of the full set of controls. An intercept of .50 is the expected value if choice is completely random and unaffected by the pitch manipulation. The coefficient on the predictors gives the deviation of the choice ratio for the group coded '+1' and the grand mean.

In both the models predicting social rank and dominance, the intercept values of .32 and .21 deviate substantially from .50 and are skewed toward 0, indicating an overall tendency to choose the deepening voice recording as more descriptive of these traits. In contrast, the intercept in the model predicting perceptions of prestige is .53, which suggests a slight bias toward choosing the *rising* pitch recording.

Wald test was used to investigate whether the intercept in each model differs from .50, which indicates random choice. As expected, the intercept in the dominance model differed significantly from .50 [$F(1, 269) = 4.48, p = .035$], whereas the intercept in the prestige model did not differ from .50 [$F(1, 269) = .07, p = .795$]. The dominance effect was thus robust to the inclusion of these subject-level controls. The intercept in the social rank model, however, was only marginally significant at conventional levels [$F(1, 269) = 2.50, p = .115$]

To further explore the effects of these controls on the social rank composite, we re-ran the same regression models but sequentially added gender, age, stimuli presentation order, and method of recruitment to a baseline model (with no predictors), and again in each case examined the estimated intercept, which provides

an unweighted grand mean of voice preference across all subjects. These regressions are supplied in Table S12, where Model 5 corresponds to the Social Rank Composite model in Table S11. As above, descriptively, the intercept across all specifications is lower than .50 and skewed towards 0, indicating that subjects tend to choose the deepening voice recording as more descriptive of rank-related traits. In fact, the inclusion of additional controls *lowers* the predicted intercept, which goes from .41 in the baseline model with no predictors to .32 in the final model that includes all 4 predictors. In Models 1 to 4, the estimation of the intercept is rather precise, with standard errors ranging from .02 to .08; these intercepts differ significantly from .50 ($ps = .0003, .0004, .018, .021$, and $.115$). However, despite a comparable effect size, the intercept in Model 5 is not well estimated due to the substantially unbalanced nature of the recruitment method control ($n = 93$ Mechanical Turk Workers; $n = 181$ undergraduate students) and the smaller sample size of the Mechanical Turk subjects. The standard error of the intercept in this model jumps up to .12 and the intercept is significant only at the 11.53% level.

Taken at face value, these additional analyses suggest that subjects perceive the deepening voice as signaling greater rank seeking, compared to the rising voice. These checks suggest that it is unlikely that this effect is driven by methodological artifacts or subject characteristics.

TABLE S11. ORDINARY LEAST SQUARE REGRESSIONS PREDICTING CHOICE OF STIMULI (AVERAGED ACROSS RANK ITEMS, DOMINANCE ITEMS, AND PRESTIGE ITEMS) FROM PARTICIPANT-LEVEL CONTROL VARIABLES. COEFFICIENTS ARE FOLLOWED BY STANDARD ERRORS IN PARENTHESES.

	Social Rank Composite	Dominance Composite	Prestige Composite
Gender (Male = -1, Female = +1)	0.01 (0.03)	0.02 (0.03)	-0.05 ⁺ (0.03)
Age	0.00 (0.00)	0.01 ⁺ (0.01)	-0.00 (0.00)
Order of Stimuli (Lowering Pitch First = -1, Raising Pitch First = +1)	0.04 (0.02)	0.06* (0.03)	-0.05 ⁺ (0.03)
Recruitment Method (Undergraduate Student = -1, Mechanical Turk Worker = +1)	0.00 (0.04)	-0.03 (0.04)	-0.02 (0.04)
(Intercept)	0.32 ^a (0.12)	0.21 ^b (0.14)	0.53 ^c (0.13)
<i>AIC</i>	270.00	367.63	320.03
<i>BIC</i>	288.06	385.70	338.09
<i>N</i>	274	274	274

⁺ $p < 0.10$, * $p < 0.05$ (for $H_0: b = 0$)

Note. Wald tests were used to test whether the model intercept differs significantly from .50 ($H_0: b_0 = .50$), which indicates random choice. A model intercept closer to 0 indicates the tendency to select the deepening voice as more characteristic of the trait, relative to the rising voice. ^a $p = .115$, ^b $p = .035$, ^c $p = .795$.

TABLE S12. ORDINARY LEAST SQUARE REGRESSIONS PREDICTING CHOICE OF STIMULI (AVERAGED ACROSS RANK ITEMS). COEFFICIENTS ARE FOLLOWED BY STANDARD ERRORS IN PARENTHESES.

	Model 1	Model 2	Model 3	Model 4	Model 5
Gender (Male = -1, Female = +1)		-0.00 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.03)
Age			0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Order of Stimuli (Lowering Pitch First = -1, Raising Pitch First = +1)				0.04 (0.02)	0.04 (0.02)
Recruitment Method (Undergraduate Student = -1, Mechanical Turk Worker = +1)					0.00 (0.04)
(Intercept)	0.41 ^a	0.41 ^b	0.31 ^c	0.31 ^d	0.32 ^e

	(0.02)	(0.02)	(0.08)	(0.08)	(0.12)
<i>AIC</i>	266.54	268.54	268.66	268.00	270.00
<i>BIC</i>	270.16	275.77	279.50	282.45	288.06
<i>N</i>	274	274	274	274	274

Note. Wald tests were used to test whether the model intercept differs significantly from .50 ($H_0: b_0 = .50$), which indicates random choice. A model intercept closer to 0 indicates the tendency to select the deepening voice as more characteristic of rank seeking, relative to the rising voice. ^a $p = .0003$, ^b $p = .0004$, ^c $p = .018$, ^d $p = .021$, ^e $p = .115$.

SUPPLEMENTAL REFERENCES

- Apicella, C. L., & Feinberg, D. R. (2009). Voice pitch alters mate-choice-relevant perception in hunter-gatherers. *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1659), 1077–1082. <http://doi.org/10.1098/rspb.2008.1542>
- Bassiri, D. (1988). Large and small sample properties of maximum likelihood estimates for hierarchical linear models. *Unpublished Doctoral Dissertation*, Michigan State University, East Lansing.
- Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, 82(1), 55–59.
- Cartei, V., Cowles, H. W., & Reby, D. (2012). Spontaneous voice gender imitation abilities in adult speakers. *PLoS ONE*, 7(2), e31353. <http://doi.org/10.1371/journal.pone.0031353>
- Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology*, 104(1), 103–125. <http://doi.org/10.1037/a0030398>

- Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, *117*(3), 319–331. <http://doi.org/10.1016/j.cognition.2010.09.003>
- Ives, D. T., Smith, D. R. R., & Patterson, R. D. (2005). Discrimination of speaker size from syllable phrases. *The Journal of the Acoustical Society of America*, *118*(6), 3816–3822. <http://doi.org/10.1121/1.2118427>
- Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, *1*(79), 57–62. <http://doi.org/10.1016/j.anbehav.2009.10.003>
- Kreft, I., & de Leeuw, J. de. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.
- Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology*, *97*(5), 951–966. <http://doi.org/10.1037/a0028380>
- Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, *14*(9), 1105–1107. <http://doi.org/10.1038/nn.2886>
- Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America*, *129*(4), 2201–2212. <http://doi.org/10.1121/1.3552866>

- Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. C. (2007). Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior, 28*(5), 340–344.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–164. <http://doi.org/10.2307/271063>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Newbury Park, CA: Sage.
- Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability the difference score in the measurement of change. *Journal of Educational Measurement, 20*(4), 335–343. <http://doi.org/10.1111/j.1745-3984.1983.tb00211.x>
- Scherbaum, C. A., & Ferrerter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods, 12*(2), 347–367. <http://doi.org/10.1177/1094428107308906>
- Sinnott, J. M., Owren, M. J., & Petersen, M. R. (1987). Auditory frequency discrimination in primates: Species differences (Cercopithecus, Macaca, Homo). *Journal of Comparative Psychology, 101*(2), 126–131. <http://doi.org/10.1037/0735-7036.101.2.126>
- Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *The Journal of the Acoustical Society of America, 117*(1), 305–318.
- Snijders, T. A. B. (2005). Power and sample size in multilevel linear models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 3, pp. 1570–1573). Chichester, England: Wiley.

Snijders, T. A. B., & Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics, 18*(3), 237–259.

<http://doi.org/10.2307/1165134>

Snijders, T. A. B., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods & Research, 22*(3), 342–363.

<http://doi.org/10.1177/0049124194022003004>

Stel, M., Dijk, E., Smith, P. K., Dijk, W. W., & Djalal, F. M. (2012). Lowering the pitch of your voice makes you feel more powerful and think more abstractly. *Social Psychological and Personality Science, 3*(4), 497–502.

van der Leeden, R., & Busing, F. M. (1994). First iteration versus final IGLS/RIGLS estimators in two-level models: A Monte Carlo study with ML3. *Leiden, the Netherlands: Leiden University*.