Evolutionary Psychology
January-March, 2016: 1–10
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1474704916630317
evp.sagepub.com
SAGE

Concordant Cues in Faces and Voices: Testing the Backup Signal Hypothesis

Harriet M. J. Smith[1], Andrew K. Dunn[1], Thom Baguley[1], and Paula C. Stacey[1]

Abstract

Information from faces and voices combines to provide multimodal signals about a person. Faces and voices may offer redundant, overlapping (backup signals), or complementary information (multiple messages). This article reports two experiments which investigated the extent to which faces and voices deliver concordant information about dimensions of fitness and quality. In Experiment 1, participants rated faces and voices on scales for masculinity/femininity, age, health, height, and weight. The results showed that people make similar judgments from faces and voices, with particularly strong correlations for masculinity/femininity, health, and height. If, as these results suggest, faces and voices constitute backup signals for various dimensions, it is hypothetically possible that people would be able to accurately match novel faces and voices for identity. However, previous investigations into novel face–voice matching offer contradictory results. In Experiment 2, participants saw a face and heard a voice and were required to decide whether the face and voice belonged to the same person. Matching accuracy was significantly above chance level, suggesting that judgments made independently from faces and voices are sufficiently similar that people can match the two. Both sets of results were analyzed using multilevel modeling and are interpreted as being consistent with the backup signal hypothesis.

Keywords

face, voice, static, dynamic, backup signal

Date received: September 18, 2015; Accepted: November 14, 2015

Together, faces and voices convey multimodal signals. Such signals are common in animals and occur when information about an underlying trait is communicated by more than one modality. As most research has focused on face and voice ratings independently of each other (Wells, Baguley, Sergeant, & Dunn, 2013; Wells, Dunn, Sergeant, & Davies, 2009), relatively little is known about multimodal signals in humans. Multimodal signals are either backup signals (Johnstone, 1997), or multiple messages (Møller & Pomiankowski, 1993), and are likely to have adaptive value in terms of mate choice. Backup signals are redundant in meaning: they offer similar information and elicit the same response, thereby helping to reduce inaccurate trait assessments (Møller & Pomiankowski, 1993). It is therefore possible to distinguish between multiple messages and backup signals by empirically testing the effect of multimodal signals on a recipient (Partan & Marler, 1999). If a multimodal signal present in human faces and voices is a backup signal for a certain dimension, ratings on this dimension should correlate, whereas uncorrelated ratings would reflect the presence of multiple messages (Wells et al., 2013; Wells et al., 2009).

Multimodal Signals in Faces and Voices

Faces and voices are salient social stimuli, offering a multitude of identity and affective information (Belin, Fecteau, & Bedard, 2004). From an evolutionary perspective, faces and voices provide valuable clues about fitness. For example, in terms of attractiveness they appear to constitute reliable and concordant signals of genetic quality (e.g., Collins & Missing, 2003; Feinberg, 2008; Feinberg et al., 2005; Fraccaro et al., 2010; Saxton, Caryl, & Roberts, 2006; Thornhill & Gangestad, 1999; Thornhill & Grammer, 1999; Wheatley et al., 2014; Zahavi & Zahavi, 1997; see also Puts, Jones, & DeBruine, 2012 for a review), and a number of studies have found that people who have faces that rate highly for attractiveness also

[1] Psychology Division, Nottingham Trent University, Nottingham, UK

Corresponding Author:
Harriet M. J. Smith, Psychology Division, Nottingham Trent University, Burton Street, Nottingham, NG1 4BU, UK.
Email: harriet.smith2011@my.ntu.ac.uk

Creative Commons CC-BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 3.0 License (http://www.creativecommons.org/licenses/by-nc/3.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).

tend to have voices that rate highly for attractiveness (e.g., Collins & Missing, 2003; Saxton et al., 2006, but see Oguchi & Kikuchi, 1997; Wells et al., 2013).

With the exception of the attractiveness literature, previous research has rarely compared judgments made from faces and voices, focusing instead on judgments informed by a single modality (e.g., Neiman & Applegate, 1990; Penton-Voak & Chen, 2004; Perrett et al., 1998; Pisanski, Mishra, & Rendall, 2012). However, there are a number of reasons as to why we may expect concordance between face and voice ratings in terms of masculinity and femininity, health, age, height, and weight. Some of these reasons are detailed below.

Masculinity/femininity. Levels of reproductive hormone levels are likely to influence perceptions of both facial and vocal femininity and masculinity. For example, testosterone increases the size and thickness of vocal folds (Beckford, Rood, & Schaid, 1985), resulting in lower fundamental frequency (Fant, 1960), which influences perceptions of masculinity (Pisanski et al., 2012). In addition, high levels of testosterone are associated with characteristics of facial masculinity (Penton-Voak & Chen, 2004; Perrett et al., 1998), such as larger jaws, chins, and noses (Miller & Todd, 1998). In women, estrogen slows down vocal fold development and is associated with higher vocal pitch (Abitbol, Abitbol, & Abitbol, 1999; O'Connor, Re, & Feinberg, 2011). Estrogen levels are also related to markers of facial femininity (Thornhill & Grammer, 1999) such as larger lips, smaller lower faces, and fat deposits on the upper cheeks (Perrett et al., 1998).

Health. We might also expect ratings of health made from faces and voices to be similar. Previous research suggests that cues relating to higher levels of reproductive hormones are reliable indicators of fitness and quality (Folstad & Karter, 1992; Thornhill & Gangestad, 2006; Zahavi & Zahavi, 1997), and, indeed, some studies suggest that measures of sexual dimorphism are linked to health ratings and actual health in both men (Gray, Berlin, McKinlay, & Longcope, 1991; Rhodes, Chan, Zebrowitz, & Simmons, 2003) and women (Ellison, 1999; Law Smith et al., 2006).

Age. Faces and voices index information about biological age, a cue which is relevant to reproductive fitness in both males and females (Thornhill & Gangestad, 1999). Numerous visual markers act as indicators of older age, such as decreased elasticity in the skin, wrinkles, discoloration, and reduced clarity in skin tone (Burt & Perrett, 1995). In terms of voices, older people speak with a slower speech rate (Linville, 1996), and age-related hormonal changes affect pitch. For example, female voice pitch lowers after the menopause, whereas older male voices become higher pitched (Linville, 1996). People can estimate a speaker's age from their voice relatively accurately (to within about 10 years; Braun, 1996; Neiman & Applegate, 1990; Ptacek & Sander, 1966; Smith & Baguley, 2014).

Height and weight. Body size is a further indicator of quality (Collins & Missing, 2003; Thornhill & Gangestad, 1999). However, although people tend to agree about height and weight judgments made from a voice (Collins, 2000), this does not indicate that they are necessarily accurate (Bruckert, Liénard, Lacroix, Kreutzer, & Leboucher, 2006; Collins, 2000; van Dommelen & Moxness, 1995). Despite the apparent inaccuracy of height judgments made from voices, people judge height from faces with relative accuracy (Schneider, Hecht, Stevanov, & Carbon, 2013), using cues such as facial elongation. People with longer faces are judged as being taller (Re et al., 2013). Judgments from faces are also accurate for weight estimates (Coetzee, Chen, Perrett, & Stephen, 2010). Lass and Colt (1980) compared visual and auditory height and weight ratings. Results showed significant differences between weight ratings from female faces and voices, suggesting that for some characteristics, faces and voices may not offer concordant information. Recent research has not addressed the extent of concordance between body size information offered by faces and voices. Although Krauss, Freyberg, and Morsella (2002) asked participants to rate the age, height, and weight of speakers from faces and voices, they only tested whether the ratings were accurate, rather than whether there was a relationship between face and voice ratings.

Static and Dynamic Faces

The extent to which faces and voices offer concordant information might be affected by whether the face is static or dynamic. For example, Lander (2008) found that male face and voice attractiveness was only related when faces were dynamic. Studies investigating facial attractiveness and human mate preferences most frequently use static facial stimuli (photos). However, there has been a recent move to use dynamic facial stimuli (videos) in order to improve ecological validity (Gangestad & Scheyd, 2005; Penton-Voak & Chang, 2008; Roberts, Saxton et al., 2009b). Some studies have found that facial stimulus type (static or dynamic) influences attractiveness judgments, although the overall results are somewhat mixed (e.g., Lander, 2008; Penton-Voak & Chang, 2008; Roberts, Little, et al., 2009a; Rubenstein, 2005). In reviewing previous studies and investigating methodological differences between them, Roberts, Saxton et al. (2009b) reported that correlations between ratings from static and dynamic facial stimuli were stronger when rated by the same participants, likely because of carryover effects. As patterns of facial movement vary according to sex (Morrison, Gralewski, Campbell, & Penton-Voak, 2007), it is conceivable that masculinity/femininity ratings will be more extreme when viewing dynamic faces. In light of these findings, it is necessary to consider the influence of facial stimulus type when testing the concordance of face–voice judgments.

Face–voice matching provides a further test of the extent to which faces and voices offer redundant information. However, it is not clear from the literature whether accurate face–voice matching using static facial stimuli is possible. While Kamachi, Hill, Lander, and Vatikiotis-Bateson (2003) showed that participants could match dynamic muted faces saying different

sentences to voices of the same identity, participants performed at chance level when the facial stimuli were static. Similar results were reported by Lachs and Pisoni (2004). However, Mavica and Barenholtz (2013) observed above chance level accuracy on trials featuring static faces, suggesting that above chance matching ability is not dependent on being able to encode visual articulatory patterns but rather on concordant information offered by faces and voices.

Aims

This article investigates the extent to which faces and voices offer concordant information, thereby providing a test of the backup signal hypothesis (Johnstone, 1997). Using both static and dynamic facial stimuli, we tested cross-modal concordance by asking participants to make judgments from faces and voices about perceived femininity/masculinity, health, age, height, and weight. In a further test of face–voice concordance, we investigated whether it is possible to accurately match novel static or dynamic faces and voices of the same identity. If faces and voices offer similar information, and it is possible to match the two, this would offer support for the backup signal hypothesis.

Experiment 1

Experiment 1 tested whether faces and voices offer concordant information about dimensions of fitness and quality, aiming to establish whether people make similar judgments about a novel person, regardless of whether they see their face or hear their voice. We expect that as the previous literature suggests that both faces and voices honestly signal quality, judgments made independently from faces and voices should be similar. In light of the contradictory findings regarding judgments made from static and dynamic facial stimuli, the study also tested whether the relationship between face and voice ratings differs according to facial stimulus type (static vs. dynamic).

Method

Design

This experiment employed a mixed design. The between-subject factor was facial stimulus type (static or dynamic), and the within-subject factor was modality (face or voice)

Participants

The participants ($n = 48$) were recruited from the Nottingham Trent University Psychology Division's Research Participation Scheme. There were 12 male and 36 female participants (age range = 18–28 years, $M = 20.54$, $SD = 2.59$). Participants gave informed consent and received a research credit in line with course requirements. The College Research Ethics Committee for Business, Law and Social Sciences granted ethical approval for the study (ref: 2013/37). All participants reported having normal to corrected hearing and vision.

Apparatus and Materials

Stimulus faces and voices were taken from the Grid audiovisual sentence corpus (Cooke, Barker, Cunningham, & Shao, 2006), a multi-talker corpus featuring head and shoulder videos of British adult speakers saying 1,000, six-word sentences each in an emotionally neutral manner recorded against a plain blue background. Each sentence follows the same six-word structure: (1) command, (2) color, (3) preposition, (4) letter, (5) digit, and (6) adverb, for example, "Place blue at J 9 now." None of the speakers in the corpus say the same sentence. A total of 18 speakers were selected from the corpus: 9 males and 9 females. Speakers were matched for ethnicity (White British), accent (English), and age (18–30).

The stimuli were presented on an Acer Aspire laptop (screen size 15.6 inches, resolution 1,366 × 768 pixels, Dolby Advanced Audio) placed approximately 8.5 cm away from the edge of the desk at which participants sat. The experiment was run using Psychopy v1.77.01 (Peirce, 2009), an open-source software package designed for running experiments in Python. Three videos (.mpegs) were selected at random from the GRID corpus for each speaker, using an online research randomizer (Urbaniak & Plous, 2013). The study used static faces, dynamic faces, and voices. One of the three videos was used to create static pictures of faces. Pictures were extracted using the snapshot function on Windows Movie Maker (2012) and presented in .png format. The static picture for each talker was the first frame of the video. Another of the three video files was used to construct the dynamic stimuli. The file was muted using Windows Movie Maker and converted back into .mpeg format. All facial stimuli measured 384 × 288 pixels and were presented in color for 2 s, with brightness settings at the maximum level. Voice recordings were also played for 2 s, from the third .mpeg file, but the face was not visible at presentation. To reduce the background noise, participants listened to the recordings binaurally through Apple earphones with a frequency range of 5–21,000 Hz. This exceeds the range of human hearing (Feinberg et al., 2005). Voices were played at a comfortable listening volume (30% of the maximum volume). Two versions of the experiment were constructed: one using static faces and voices and the other using dynamic faces and voices. In both versions, all 18 faces and voices appeared.

Procedure. Participants were randomly allocated to either the static face or the dynamic face version of the experiment. They read the information sheet, completed the consent form, and provided demographic information. Testing took place in a quiet cubicle. Participants completed two counterbalanced blocks of testing. In one block participants viewed faces, in the other they heard voices. Participants were not told that the voices and faces featured in the experiment belonged to the same people. Each block consisted of a practice trial followed by 18 randomly ordered experimental trials. After each face or voice, participants estimated the age of the stimulus person in years and completed the 7-point Likert-style rating scales in the following order: femininity/masculinity (1 = *very feminine*,

7 = *very masculine*), health (1 = *very unhealthy*, 7 = *very healthy*), height (1 = *very short*, 7 = *very tall*), and weight (1 = *very underweight*, 7 = *very overweight*).

Data Analysis and Multilevel Modeling

Data were analyzed using multilevel models, rather than performing conventional analyses on data averaged over either participants or stimuli (see Wells et al., 2013). This avoids the ecological fallacy which arises when it is falsely assumed that patterns observed for participant means also hold for data at a lower level of analysis such as individual trials repeated within participants (e.g., see Robinson, 1950; Wells et al., 2013). Multilevel modeling allows both participants and stimuli to be simultaneously treated as random effects, thereby maximizing generalizability (Clark, 1973; Judd, Westfall, & Kenny, 2012). When the random effects are fully crossed (i.e., when all participants experience all stimuli), conventional analyses (including separate by-items or by-subjects analyses) can lead to massive Type 1 error inflation (Baguley, 2012; Clark, 1973; Judd et al., 2012). The most appropriate analysis therefore takes into account both sources of variability. Unless the ignored source of variability is negligible, this is always more conservative than separate by-stimuli or by-participants analyses.

Results

We calculated the absolute difference between face and voice ratings by comparing each rating participants had given to a face and voice belonging to the same person. Then we calculated the mean absolute difference (MAD) for each stimuli person on each rating scale (age, masculinity/femininity, health, height, and weight). Descriptive statistics (Table 1) indicate that typical ratings for faces and voices fall within a similar range.

On all scales apart from age, face and voice ratings only differ on average by about 1 point (14%) on a 7-point rating scale, and MADs were similar across static and dynamic facial stimuli. The difference between face and voice ratings in terms of age appears larger than that of the other rating scales. However, rather than being rated on a 7-point scale, age estimates were given in years. This prevents a neat comparison between the rating scales.

The results in Table 1 show that face and voice ratings tend to be close together in terms of the range they fall into. A logical next step is to quantify the extent to which voice and face ratings covary in the same individual. For this purpose, a simple correlation coefficient between voice and face ratings would either ignore the dependency within participants or rely only on aggregate data (mean ratings for each participant). We therefore used multilevel models to account for both participant and stimuli variation when correlating voice ratings with face ratings for estimated age and ratings for femininity/masculinity, health, height, and weight. For each variable, we fitted an intercept-only model with the rating as an outcome, using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2014). A crucial part of each model was to estimate separate variance for face and voice ratings as well as the correlation between face and voice ratings across both stimuli and participants. The correlation between face and voice ratings within participants is, for present purposes, a nuisance term (merely indicating that participants who give high ratings to voices also tend to give high ratings to faces) and is not reported here. The correlations reported in Table 2 are those within stimuli and demonstrate that, for a given item, voice and face ratings are positively correlated.

Table 1. MAD and 95% Confidence Intervals for the MAD Between Face and Voice Ratings by Stimulus-Type Condition.

	Static Facial Stimuli				Dynamic Facial Stimuli			
			95% CI				95% CI	
Rating scale	*M*	*SD*	LB	UB	*M*	*SD*	LB	UB
Age	3.91	1.51	3.27	4.55	3.62	1.58	2.95	4.29
Masculinity/femininity	1.05	0.35	0.90	1.19	1.00	.36	0.85	1.15
Health	1.24	.34	1.10	1.39	1.12	0.27	1.00	1.23
Height	1.10	.29	0.98	1.23	1.04	0.36	0.89	1.19
Weight	0.92	0.25	0.81	1.02	1.00	0.27	0.88	1.11

Note. MAD = mean absolute difference.

Table 2. Within-Stimulus Correlations Between Face and Voice Ratings.

	Correlation coefficient				
Condition	Age	Masc/fem	Health	Height	Weight
Static facial stimuli	.60	.97	.70	.83	.40
Dynamic facial stimuli	.32	.92	.91	.86	.17
All facial stimuli	.46	.95	.77	.84	.28

Table 2 provides evidence that mean face and voice ratings for the same target appear to be positively related for all rating types. Correlations between face and voice ratings on scales for masculinity/femininity, health, and height were particularly high, regardless of whether the facial stimuli were static or dynamic. Correlations between mean face and voice ratings for age and weight were moderate when facial stimuli were static—with some suggestion that the correlations were diminished for dynamic stimuli. However, correlations did not vary according to facial stimulus type in direction or by more than .3 on any scale. The difference between the static and dynamic correlations was tested by fitting models with separate variance terms for each stimulus type. Comparing a model which includes separate variance and covariance terms for static and dynamic stimuli with one that does not did not improve the model fit for any of the ratings ($p > .14$). This complements the results shown in Table 1, suggesting that the extent to which faces and voices offer similar information is not greatly influenced by whether the facial stimuli is static or dynamic.

Discussion

Experiment 1 showed that observers glean concordant information about different dimensions of quality from faces and

voices, particularly in terms of masculinity and femininity, health, and height. On each dimension, the relatedness of face and voice ratings is not affected by facial stimulus type, showing that the signals tested here are stable across static and dynamic faces. These results support the hypothesis that on various dimensions of quality, faces and voices constitute backup signals.

Experiment 2

Experiment 2 tested whether faces and voices offer sufficiently concordant information that people can match novel faces to voices. Previous studies have addressed this question, with conflicting results. Krauss et al. (2002) showed that people are relatively accurate at inferring physical information from a voice. After only hearing a voice excerpt, participants selected the speaker's full-length photograph from one of two possible options with above chance accuracy. Mavica and Barenholtz (2013) tested whether people could use information from a voice to distinguish between two static images of different faces. Accuracy was significantly above chance level, despite contradictory results presented in previous studies (Kamachi et al., 2003; Lachs & Pisoni, 2004) suggesting that successful matching of faces and voices depends on the ability to encode dynamic properties of speaking (muted) faces (Mavica & Barenholtz, 2013).

Previous face–voice matching studies (Kamachi et al., 2003; Krauss et.al., 2002; Mavica & Barenholtz, 2013) have used a two-alternative forced choice paradigm (2AFC), which unlike a same–different paradigm does not model whether people are also able to correctly reject a match when a face and voice are from different people. The 2AFC tasks therefore give no information about possible response biases. Experiment 2 uses a same–different paradigm to give a clearer picture of face–voice matching ability.

Experiment 2 addresses three main questions. First, whether it is possible to accurately match novel faces and voices of the same age (20–30), sex, and ethnicity (White British). Second, whether matching accuracy is affected by facial stimulus type (static or dynamic). Third, in line with cross-modal matching procedures (Kamachi et al., 2003; Lachs & Pisoni, 2004), we investigated whether people are more accurate at face–voice matching when visual information (a face) is presented first, compared to when auditory information (a voice) is presented first. If faces and voices primarily constitute backup signals, people should be able to match novel faces and voices above chance level.

Method

The methods for Experiment 2 were the same as for Experiment 1, with exceptions explained in the following subsections.

Design

This experiment employed a $2 \times 2 \times 2$ mixed factorial design. The between-subject factor was facial stimulus type (static or dynamic). The within-subject factors were identity (same or different) and order (face first or voice first). The dependent variable was accuracy.

Participants

There were 40 male and 40 female adult participants ($n = 80$) with an age range of 18–66 years ($M = 25.44$, $SD = 8.36$).

Materials

Four different versions of the experiment were created so that matching and not-matching pairs of faces and voices could be constructed using different stimulus people. Stimuli were randomly selected to be used for either one of the eight same identity or eight different identity trials. None of the faces or voices appeared more than once in each version. On different identity trials, the face and voice were matched for age, gender, and ethnicity. The stimuli that remained were used for the practice trials. Each version was repeated for static and dynamic conditions. In total, there were eight versions.

Procedure

Participants were randomly allocated to one of the eight versions of the experiment. In the dynamic facial stimulus condition, participants were also correctly informed that the face in the muted video and the voice in the recording were not saying the same thing. This was to prevent them using speech reading to match the face and voice (Kamachi et al., 2003).

Participants completed two counterbalanced experimental blocks, each consisting of a practice trial followed by eight randomly ordered experimental trials. In one block, participants saw the face first, and in the other they heard the voice first. None of the stimuli appeared more than once in each version of the experiment. In each trial, there was a 1-s gap between presentation of the face and voice stimuli. At test, participants pressed "1" if they thought the face and voice were "matching" (same identity), and "0" if they thought it was "not matching" (different identity).

Results

Performance accuracy was analyzed using multilevel logistic regression with the lme4 version 1.06 package in R (Bates et al., 2014). Four nested models with accuracy (0 or 1) as the dependent variable were compared (and all models were fitted using restricted maximum likelihood). The first model included a single intercept (and was later used to obtain confidence intervals for the overall accuracy). The second model also included the main effects of each factor (identity, order, and stimulus type). The third model added all two-way interactions and the final model added the three-way interaction. Setting up the model in this way allows us to test for individual effects in a manner similar to that of a traditional analysis of variance. However, as F-tests-derived multilevel models are not, in general, accurate, we report the more robust profile likelihood ratio

Table 3. Parameter Estimates (*b*) and Profile Likelihood Tests for the 2 × 2 × 2 Factorial Analysis of Accuracy in Experiment 2.

Source	*df*	*b*	*SE*	G^2	*p*
Intercept	1	−0.445	0.196		
Identity	1	1.382	0.254	57.84	<.001
Order	1	0.509	0.241	2.28	.131
Facial stimulus type	1	0.133	0.231	0.13	.717
Identity × Order	1	0.601	0.358	4.20	.040
Identity × Facial Stimulus Type	1	0.165	0.339	0.32	.572
Order × Facial Stimulus Type	1	0.052	0.324	0.01	.916
Identity × Order × Facial Stimulus Type	1	0.058	0.474	0.01	.903

tests provided by lme4. These were obtained by dropping each effect in turn from the appropriate model (e.g., testing the three-way interaction by dropping it from the model including all effects, and testing the two-way interactions by dropping each effect in turn from the two-way model).

Table 3 shows the profile likelihood chi-square statistic (G^2) and *p*-value associated with dropping each effect. Table 3 also reports the coefficients and standard errors (on a log odds scale) for each effect in the full three-way interaction model. In the three-way model, the estimate of *SD* of the face random effect was 0.353, while for voice it was 0.207. The estimated *SD* for the participant effect was less than 0.0001. A similar pattern held for the null model. Thus, although individual differences were negligible in this instance, a conventional by-participants analysis that did not incorporate both voice and face variation could be extremely misleading.

Only the main effect of identity and the two-way interaction of identity and order were statistically significant. To aid interpretation of these effects, we obtained means and confidence intervals for the percentage accuracy of the eight conditions in the factorial design. These confidence intervals were obtained through simulations of the posterior distributions of the cell means using arm package version 1.6 in R (Gelman & Su, 2013). These means and the associated 95% confidence intervals are shown in Figure 1.

From Figure 1 it is clear that overall matching performance was significantly above chance (50%) level, $M = 59.7\%$, 95% CI [51.9, 66.9]. Static face–voice matching was above chance, $M = 59.19$, 95% CI [50.94, 66.84], as was dynamic face–voice matching, $M = 60.12$, 95% CI [51.97, 67.74]. Figure 1 also reveals the main effect of identity, with performance for same trials consistently higher than for different trials (and the former but not the latter consistently above chance). It also reveals the basis of the identity by order interaction. The results from the face first trials are shown in Panel A. The results from the voice first trials are shown in Panel B. Although same identity trials showed better performance than different trials for both face first and voice first trials, this advantage is greater in the face first conditions. Given that performance on the face first different trials is on average worse than chance (and significantly so for the static stimuli), this pattern suggests the operation of a response bias, such that participants exhibited a bias

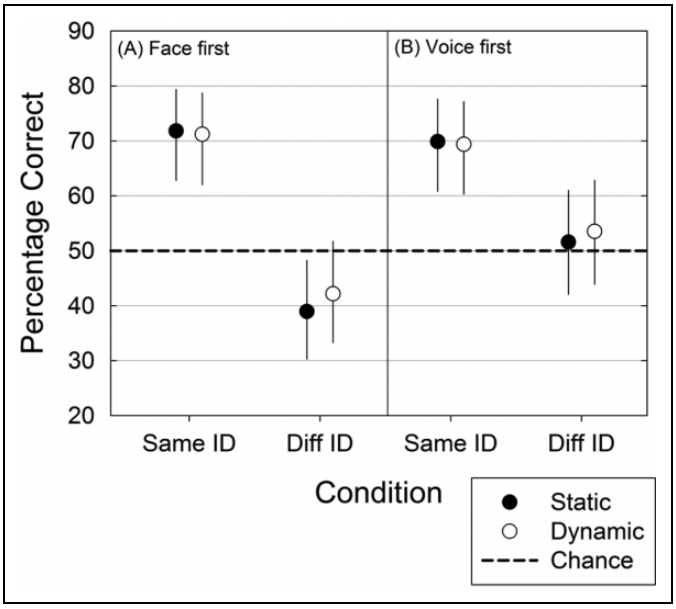

Figure 1. Face–voice matching accuracy on face first (Panel A) and voice first (Panel B) trials. Error bars show 95% CI for the condition means. CI = confidence interval.

to accept faces and voices as belonging to the same identity when they saw the face before hearing the voice.

Discussion

In Experiment 2, we observed that both dynamic faces and voices, and static faces and voices, can be matched for identity above chance level. These results are consistent with the hypotheses informed by the results of Experiment 1, which show that faces and voices offer a high level of concordant information on various dimensions. Face–voice matching performance does not differ according to facial stimulus type. Therefore, accuracy does not appear to depend on encoding visual information about speaking style but rather on redundant signals available in voices and static faces.

General Discussion

The results of Experiment 1 are consistent with the hypothesis that faces and voices offer redundant signals for various dimensions of quality. Mean face and voice ratings for the same target were positively related for all rating types. Correlations between face and voice ratings on scales for masculinity/femininity, health, and height were particularly strong, regardless of whether the facial stimuli were static or dynamic. The results of Experiment 2 show that the information signaled by faces and voices is so similar that people can match novel faces and voices of the same sex, ethnicity, and age-group at a level significantly above chance. Taken together, results suggest that faces and voices constitute backup signals, reinforcing the same information about quality (Johnstone, 1997) rather than

complementary but different information (Møller & Pomiankowski, 1993).

Face and Voice Ratings

With the exception of the attractiveness literature, previous research has rarely compared judgments made from faces and voices, focusing instead on judgments informed by a single modality (e.g., Penton-Voak & Chen 2004; Perrett et al., 1998; Pisanski et al., 2012; Neiman & Applegate, 1990, and so on) or comparing face and voice ratings to actual measurements of physical characteristics (e.g., Krauss et al., 2002) rather than to each other. The results of Experiment 1 show that not only do face and voice ratings fall within a small range but independent ratings of an individual's face and voice are positively correlated. These results complement other studies, showing that faces and voices offer related information about fitness and mate value (Collins & Missing, 2003; Feinberg, 2008; Feinberg et al., 2005; Fraccaro et al., 2010).

The strongest correlations between face and voice ratings occurred on scales for masculinity/femininity, health, and height. Despite the previous literature suggesting that unimodal voice ratings of body size are less accurate than unimodal face ratings (Bruckert et al., 2006; Coetzee et al., 2010; Collins, 2000; Re et al., 2013; van Dommelen & Moxness, 1995), Experiment 1 showed that regardless of accuracy, the MAD between body size judgments made from faces and voices was small. However, correlations were strong for height but only weak-moderate for weight. This corresponds with Lass and Colt (1980) who found significant differences between weight ratings for female faces and voices.

Face and Voice Matching

Overall, face–voice matching accuracy in Experiment 2 was significantly above chance. This result is consistent with previous findings (Krauss et al., 2002; Mavica & Barenholtz, 2013) and shows that people can use redundant information to match faces and voices of the same identity. Furthermore, the use of multilevel modeling allows us to generalize these findings beyond the sample of faces and voices used, thereby overcoming a common limitation of previous studies.

Although overall matching accuracy is at 59.7%, there is still a substantial proportion of unexplained variance which could be due to the existence of discordant rather than concordant face–voice information. Beyond the characteristics tested in Experiment 1, faces and voices also convey a multitude of other information, including personality characteristics and emotion (Belin et al., 2004; Mavica & Barenholtz, 2013), some of which might be complementary. Nevertheless, the results from Experiment 2 suggest that on balance, faces and voices provide concordant information because overall performance is significantly above chance level. These results are consistent with the results presented in Experiment 1.

On different identity trials, participants performed at chance level (voice first trials), or below chance level (face first trials), and were significantly less accurate than on same identity trials. This indicates that participants were better at detecting a correct match than rejecting an incorrect one. In line with the argument presented above, based purely on the findings from Experiment 1, we might have expected that accurately rejecting mismatches would be possible because the ratings were so closely related. It seems that participants are using other information to inform their matching decisions on different identity trials. On the other hand, the pattern of results across same–different trials might be partially explained by the existence of a response bias.

While previous face–voice matching studies using 2AFC procedures have found no difference between face first and voice first performance (Kamachi et al., 2003; Lachs & Pisoni, 2004), our results using a same–different task suggest people exhibit a bias to respond that a face and voice belong to the same identity, particularly when the face is presented before the voice. A performance asymmetry, according to stimuli order, is consistent with the previous literature. For instance, studies have consistently found asymmetries between faces and voices in terms of rates of recognition accuracy, which have been attributed to differential link strength in the two perception pathways (e.g., Damjanovic & Hanley, 2007; Hanley & Turner, 2000; Stevenage, Hugill, & Lewis, 2012). Therefore, there is no reason to assume that face first and voice first matching performance should be identical. However, based on the finding that familiar faces prime familiar voices better than familiar voices prime familiar faces (Stevenage et al., 2012), we might have expected the asymmetry to operate the other way around. Nevertheless, it is feasible that voices give more information about faces than faces do about voices, and aside from conveying semantic information about the spoken message, the other important role of voices is to allow people to infer socially relevant visual information about the speaker, such as information about masculinity/femininity, body size, health, and age. This idea is in keeping with the finding that showing participants mismatched celebrity face–voice pairs disrupts voice recognition to a greater extent than it disrupts face recognition (Stevenage, Neil, & Hamlin, 2014). During social interactions, it is common to hear a voice while not looking in the direction of the speaker. Being able to accept or reject a face match quickly may aid social communication by facilitating attention shifts.

Static and Dynamic Faces

Informed by contradictory findings relating to the effect of static and dynamic facial stimuli on ratings of attractiveness (e.g., Lander, 2008; Roberts, Little, et al., 2009a; Rubenstein, 2005) and face–voice matching ability (Kamachi et al., 2003; Lachs & Pisoni, 2004; Mavica & Barenholtz, 2013), we tested whether facial stimulus type affected the extent of face–voice concordance. In both experiments, performance was unaffected by whether the facial stimuli were dynamic or static. This suggests that information on these dimensions is stable across dynamic and static faces. Novel face–voice matching ability is not due to encoding visual articulatory patterns (Mavica & Barenholtz, 2013) but to the availability of redundant information.

Stimulus Sample Size

The findings of the multilevel models we report emphasize the importance of stimulus sample size in estimating effects. These models provide the tools to generalize over both participants and stimuli, but obtaining large samples of stimuli is challenging. The corpus (Cooke et al., 2006) we used only contained 18 stimulus individuals matched for age, gender, and ethnicity. This reduced the set of stimuli available for study but also reduced extraneous variability. In addition, all of the people in this stimulus set were from similar educational backgrounds (Cooke et al., 2006), and none of them exhibited strong regional accents. As there is a high level of interstimulus variability in both faces (Valentine, Lewis, & Hills, 2015) and voices (Stevenage & Neil, 2014), we would encourage future face–voice matching studies to aim for larger samples of stimuli, having demonstrated that it is variation in faces and voices that is the limiting factor on statistical power in experiments such as these (as face and voice variation is consistently higher than participant variation). However, many published studies have used samples of stimuli far smaller than 18 when investigating person perception (see G. L. Wells & Windshitl, 1999), as have other face–voice matching studies (e.g., Lachs & Pisoni, 2004). Crucially, only by accounting for variability in stimuli is it reasonable to generalize from stimuli as well as participants. Even in studies using large sample of stimuli, generalizability is limited by the common practice of aggregating over stimuli (Clark, 1973; Judd et al., 2012; Wells et al., 2013). Ultimately, the adequate sample size of stimuli or participants in experimental designs such as those reported here is a question of statistical power (e.g., see Westfall, Kenny, & Judd, 2014).

Conclusion

Faces and voices of the same identity offer redundant signals about a number of dimensions associated with quality and fitness. Information about masculinity/femininity, height, and health is particularly similar across faces and voices. We have shown that the level of redundancy between faces and voices is sufficient that it is possible to accurately match them for identity. In summary, the results of Experiments 1 and 2 are more consistent with the backup signal hypothesis (Johnstone, 1997) than the multiple messages hypothesis (Møller & Pomiankowski, 1993). As multimodal signals for various indicators of quality, faces, and voices offer concordant rather than complementary information.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The first author was supported by a Ph.D. studentship from the Division of Psychology, Nottingham Trent University.

References

Abitbol, J., Abitbol, P., & Abitbol, B. (1999). Sex hormones and the female voice. *Journal of Voice*, *13*, 424–446. doi:10.1016/S0892-1997(99)80048-4

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods*, *44*, 158–175. doi:10.3758/s13428-011-0123-7

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. Retrieved September 12, 2014, from http://CRAN.R-project.org/package=lme4

Beckford, N. S., Rood, S. R., & Schaid, D. (1985). Androgen stimulation and laryngeal development. *Annals of Otology, Rhinology, and Laryngology*, *94*, 634–640.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*, 129–135. doi:10.1016/j.tics.2004.01.008

Braun, A. (1996). Age estimation by different listener groups. *International Journal of Speech Language and the Law*, *3*, 65–73. doi:10.1558/ijsll.v3i1.65

Bruckert, L., Liénard, J. S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society, B: Biological Sciences*, *273*, 83–89. doi:10.1098/rspb.2005.3265

Burt, D. M., & Perrett, D. I. (1995). Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society, B: Biological Sciences*, *259*, 137–143. doi:10.1098/rspb.1995.0021

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359. doi:10.1016/S0022-5371(73)80014-3

Coetzee, V., Chen, J., Perrett, D. I., & Stephen, I. D. (2010). Deciphering faces: Quantifiable visual cues to weight. *Perception*, *39*, 51–61. doi:10.1068/p6560

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour*, *60*, 773–780. doi:10.1006/anbe.2000.1523

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour*, *65*, 997–1004. doi:10.1006/anbe.2003.2123

Cooke, M., Barker, J., Cunningham, S., & Shao, X. (2006). An audio-visual corpus for speech perception and automatic speech recognition. *The Journal of the Acoustical Society of America*, *120*, 2421–2424. doi:10.1121/1.2229005

Damjanovic, L., & Hanley, J. R. (2007). Recalling episodic and semantic information about famous faces and voices. *Memory & Cognition*, *35*, 1205–1210. doi:10.3758/BF03193594

Ellison, P. T. (1999). Reproductive ecology and reproductive cancers. In C. Pater-Brick & C. Worthman (Eds.), *Hormones, health, and behavior: A socio-ecological and lifespan perspective* (pp. 184–209). Cambridge, England: Cambridge University Press.

Fant, G. (1960). *The acoustic theory of speech production*. The Hague, the Netherlands: Mouton.

Feinberg, D. R. (2008). Are human faces and voices ornaments signaling common underlying cues to mate value? *Evolutionary*

Anthropology: Issues, News, and Reviews, *17*, 112–118. doi:10.1002/evan.20166

Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Law Smith, M. J., Cornwell, R. E., . . . Perrett, D. I. (2005). The voice and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, *26*, 398–408. doi:10.1016/j.evolhumbehav.2005.04.001

Folstad, I., & Karter, A. J. (1992). Parasites, bright males, and the immunocompetence handicap. *American Naturalist*, *139*, 603–622. doi:10.1086/285346

Fraccaro, P. J., Feinberg, D. R., DeBruine, L. M., Little, A. C., Watkins, C. D., & Jones, B. C. (2010). Correlated male preferences for femininity in female faces and voices. *Evolutionary Psychology*, *8*, 447–461. doi:10.1177/147470491000800311

Gangestad, S. W., & Scheyd, G. J. (2005). The evolution of human physical attractiveness. *Annual Review of Anthropology*, *34*, 523–548. doi:10.1146/annurev.anthro.33.070203.143733

Gelman, A. E., & Su, Y. S. (2013). arm: Data analysis using regression and multilevel/hierarchical models. R package version 1.6-05. Retrieved September 12, 2014, from http://CRAN.R-project.org/package=arm

Gray, A., Berlin, J. A., McKinlay, J. B., & Longcope, C. (1991). An examination of research design effects on the association of testosterone and male aging: Results of a meta-analysis. *Journal of Clinical Epidemiology*, *44*, 671–684. doi:10.1016/0895-4356(91)90028-8

Hanley, J. R., & Turner, J. M. (2000). Why are familiar-only experiences more frequent for voices than for faces? *The Quarterly Journal of Experimental Psychology: Section A*, *53*, 1105–1116. doi:10.1080/713755942

Johnstone, R. A. (1997). The evolution of animal signals. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 155–178). Oxford, England: Blackwell.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*, 54. doi:10.1037/a0028347

Kamachi, M., Hill, H., Lander, K., & Vatikiotis-Bateson, E. (2003). Putting the face to the voice: Matching identity across modality. *Current Biology*, *13*, 1709–1714. doi:10.1016/j.cub.2003.09.005

Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology*, *38*, 618–625. doi:10.1016/S0022-1031(02)00510-3

Lachs, L., & Pisoni, D. B. (2004). Crossmodal source identification in speech perception. *Ecological Psychology*, *16*, 159–187. doi:10.1207/s15326969eco1603_1

Lander, K. (2008). Relating visual and vocal attractiveness for moving and static faces. *Animal Behaviour*, *75*, 817–822. doi:10.1016/j.anbehav.2007.07.001

Lass, N. J., & Colt, E. G. (1980). A comparative study of the effect of visual and auditory cues on speaker height and weight identification. *Journal of Phonetics*, *8*, 277–285.

Law Smith, M. J., Perrett, D. I., Jones, B. C., Cornwell, R. E., Moore, F. R., Feinberg, D. R., . . . Hillier, S. G. (2006). Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society B: Biological Sciences*, *273*, 135–140. doi:10.1098/rspb.2005.3296

Linville, S. E. (1996). The sound of senescence. *Journal of Voice*, *10*, 190–200. doi:10.1016/S0892-1997(96)80046-4

Mavica, L. W., & Barenholtz, E. (2013). Matching voice and face identity from static images. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 307–312. doi:10.1037/a0030945

Miller, G. F., & Todd, P. M. (1998). Mate choice turns cognitive. *Trends in Cognitive Sciences*, *2*, 190–198. doi:10.1016/S1364-6613(98)01169-3

Møller, A. P., & Pomiankowski, A. (1993). Why have birds got multiple sexual ornaments? *Behavioral Ecology and Sociobiology*, *32*, 167–176. doi:10.1007/BF00173774

Morrison, E. R., Gralewski, L., Campbell, N., & Penton-Voak, I. S. (2007). Facial movement varies by sex and is related to attractiveness. *Evolution and Human Behavior*, *28*, 186–192. doi:10.1016/j.evolhumbehav.2007.01.001

Neiman, G. S., & Applegate, J. A. (1990). Accuracy of listener judgments of perceived age relative to chronological age in adults. *Folia Phoniatrica et Logopaedica*, *42*, 327–330. doi:10.1159/000266090

O'Connor, J. J., Re, D. E., & Feinberg, D. R. (2011). Voice pitch influences perceptions of sexual infidelity. *Evolutionary Psychology*, *9*, 64–78. doi:10.1177/147470491100900109

Oguchi, T., & Kikuchi, H. (1997). Voice and interpersonal attraction. *Japanese Psychological Research*, *39*, 56–61. doi:10.1111/1468-5884.00037

Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, *283*, 1272–1273. doi:10.1126/science.283.5406.1272

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 1–8. doi:10.3389/neuro.11.010.2008

Penton-Voak, I., & Chang, H. (2008). Attractiveness judgements of individuals vary across emotional expression and movement conditions. *Journal of Evolutionary Psychology*, *6*, 89–100. doi:10.1556/JEP.2008.1011

Penton-Voak, I. S., & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behavior*, *25*, 229–241. doi:10.1016/j.evolhumbehav.2004.04.003

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., . . . Akamatsu, S. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature*, *394*, 884–887. doi:10.1038/29772

Pisanski, K., Mishra, S., & Rendall, D. (2012). The evolved psychology of voice: Evaluating interrelationships in listeners' assessments of the size, masculinity, and attractiveness of unseen speakers. *Evolution and Human Behavior*, *33*, 509–519. doi:10.1016/j.evolhumbehav.2012.01.004

Ptacek, P. H., & Sander, E. K. (1966). Age recognition from voice. *Journal of Speech & Hearing Research*, *9*, 273–277. doi:10.1044/jshr.0902.273

Puts, D. A., Jones, B. C., & DeBruine, L. M. (2012). Sexual selection on human faces and voices. *Journal of Sex Research*, *49*, 227–243. doi:10.1080/00224499.2012.658924

Re, D. E., Hunter, D. W., Coetzee, V., Tiddeman, B. P., Xiao, D., DeBruine, L. M., . . . Perrett, D. I. (2013). Looking like a leader–facial shape predicts perceived height and leadership ability. *PloS one*, *8*, e80957. doi:10.1371/journal.pone.0080957

Rhodes, G., Chan, J., Zebrowitz, L. A., & Simmons, L. W. (2003). Does sexual dimorphism in human faces signal health? *Proceedings of the Royal Society of London B: Biological Sciences*, *270*, S93–S95. doi:10.1098/rsbl.2003.0023

Roberts, S. C., Little, A. C., Lyndon, A., Roberts, J., Havlicek, J., & Wright, R. L. (2009a). Manipulation of body odour alters men's self-confidence and judgements of their visual attractiveness by women. *International Journal of Cosmetic Science*, *31*, 47–54. doi:10.1111/j.1468-2494.2008.00477.x

Roberts, S. C., Saxton, T. K., Murray, A. K., Burriss, R. P., Rowland, H. M., & Little, A. C. (2009b). Static and dynamic facial images cue similar attractiveness judgements. *Ethology*, *115*, 588–595. doi:10.1556/JEP.7.2009.1.4.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351–357. doi:10.2307/2087176

Rubenstein, A. J. (2005). Variation in perceived attractiveness: Differences between dynamic and static faces. *Psychological Science*, *16*, 759–762. doi:10.1111/j.1467-9280.2005.01610.x

Saxton, T. K., Caryl, P. G., & Roberts, C. S. (2006). Vocal and facial attractiveness judgments of children, adolescents and adults: The ontogeny of mate choice. *Ethology*, *112*, 1179–1185. doi:10.1111/j.1439-0310.2006.01278.x

Schneider, T. M., Hecht, H., Stevanov, J., & Carbon, C. C. (2013). Cross-ethnic assessment of body weight and height on the basis of faces. *Personality and Individual Differences*, *55*, 356–360. doi:10.1016/j.paid.2013.03.022

Smith, H. M. J., & Baguley, T. (2014). Unfamiliar voice identification: Effect of post-event information on accuracy and voice ratings. *Journal of European Psychology Students*, *5*, 59–68. doi:10.5334/jeps.bs

Stevenage, S. V., Hugill, A. R., & Lewis, H. G. (2012). Integrating voice recognition into models of person perception. *Journal of Cognitive Psychology*, *24*, 409–419. doi:10.1080/20445911.2011.642859

Stevenage, S. V., & Neil, G. J. (2014). Hearing faces and seeing voices: The integration and interaction of face and voice processing. *Psychologica Belgica*, *54*, 266–281. doi:10.5334/pb.ar

Stevenage, S. V., Neil, G. J., & Hamlin, I. (2014). When the face fits: Recognition of celebrities from matching and mismatching faces and voices. *Memory*, *22*, 284–294. doi:10.1080/09658211.2013.781654

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences*, *3*, 452–460. doi:10.1016/S1364-6613(99)01403-5

Thornhill, R., & Gangestad, S. W. (2006). Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evolution and Human Behavior*, *27*, 131–144. doi:10.1016/j.evolhumbehav.2005.06.001

Thornhill, R., & Grammer, K. (1999). The body and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, *20*, 105–120. doi:10.1016/S1090-5138(98)00044-0

Urbaniak, G. C., & Plous, S. (2013). *Research randomizer* (Version 4.0) [Computer software]. Retrieved from http://www.randomizer.org/

Valentine, T., Lewis, M. B., & Hills, P. J. (2015). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, 1–24. doi:10.1080/17470218.2014.990392

van Dommelen, W. A., & Moxness, B. H. (1995). Acoustic parameters in speaker height and weight identification: Sex-specific behaviour. *Language and Speech*, *38*, 267–287. doi:10.1177/002383099503800304

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*, 1115–1125. doi:10.1177/01461672992512005

Wells, T., Baguley, T. S., Sergeant, M. J. T., & Dunn, A. K. (2013). Perceptions of human attractiveness comprising face and voice cues. *Archives of Sexual Behavior*, *42*, 805–811. doi:10.1007/s10508-012-0054-0

Wells, T., Dunn, A. K., Sergeant, M. J. T., & Davies, M. N. O. (2009). Multiple signals in human mate selection: A review and framework for integrating facial and vocal signals. *Journal of Evolutionary Psychology*, *7*, 111–139. doi:10.1556/JEP.7.2009.2.2

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*, 2020–2045. doi:10.1037/xge0000014

Wheatley, J. R., Apicella, C. A., Burriss, R. P., Cárdenas, R. A., Bailey, D. H., Welling, L. L., & Puts, D. A. (2014). Women's faces and voices are cues to reproductive potential in industrial and forager societies. *Evolution and Human Behavior*, *35*, 264–271. doi:10.1016/j.evolhumbehav.2014.02.006

Zahavi, A., & Zahavi, A. (1997). *The handicap principle*. New York, NY: Oxford University Press.