the british
psychological society
promoting excellence in psychology

Perceptual prioritization of self-associated voices

Bryony Payne[1]*, Nadine Lavan[1,2], Sarah Knight[1,3] and
Carolyn McGettigan[1]

[1]Department of Speech, Hearing and Phonetic Sciences, University College London, UK
[2]Department of Psychology, Queen Mary University of London, UK
[3]Department of Psychology, University of York, UK

Information associated with the self is prioritized relative to information associated with others and is therefore processed more quickly and accurately. Across three experiments, we examined whether a new externally-generated voice could become associated with the self and thus be prioritized in perception. In the first experiment, participants learned associations between three unfamiliar voices and three identities (self, friend, stranger). Participants then made speeded judgements of whether voice-identity pairs were correctly matched, or not. A clear self-prioritization effect was found, with participants showing quicker and more accurate responses to the newly self-associated voice relative to either the friend- or stranger- voice. In two further experiments, we tested whether this prioritization effect increased if the self-voice was gender-matched to the identity of the participant (Experiment 2) or if the self-voice was chosen by the participant (Experiment 3). Gender-matching did not significantly influence prioritization; the self-voice was similarly prioritized when it matched the gender identity of the listener as when it did not. However, we observed that choosing the self-voice did interact with prioritization (Experiment 3); the self-voice became more prominent, via lesser prioritization of the other identities, when the self-voice was chosen relative to when it was not. Our findings have implications for the design and selection of individuated synthetic voices used for assistive communication devices, suggesting that agency in choosing a new vocal identity may modulate the distinctiveness of that voice relative to others.

Voice is central to the dynamic construction of the self and is therefore of great personal and social importance. A speaker's unique voice pattern conveys a wealth of information about their physical self to a listener, including their age, gender, health, and affective state (Kreiman & Sidtis, 2011). A speaker also has extensive control over the dynamic use of their vocal apparatus and is able to flexibly and deliberately modulate the acoustics of their voice (Hughes, Mogilski, & Harrison, 2014; McGettigan et al., 2013; McGettigan & Scott, 2014) to fulfil diverse communicative goals according to changes in social demands and the acoustic environment.

*Correspondence should be addressed to Bryony Payne, Department of Speech, Hearing and Phonetic Sciences, University College London, London, WC1N 1PF, UK (email: bryony.payne.16@ucl.ac.uk).

There is evidence that speakers are implicitly biased towards their own voice, such that they evaluate it as more attractive than the voices of others (Hughes & Harrison, 2013). This bias appears in spite of the typically negative reaction that occurs when a speaker is aware of hearing their voice on a recording (Holzman & Rousey, 1966), which suggests that the bias arises from an implicit egotism and, by extension, a preference for their voice as a self-associated and self-generated signal. It remains unclear, however, whether vocal stimuli need to be self-generated in order to elicit a self-bias. This study therefore asks whether an externally-generated vocal stimulus, such as another's voice, can become associated with the self and, in turn, elicit a processing bias.

This question is of direct relevance to patients with vocal impairments who cannot use their own vocal apparatus to produce sound but instead come to rely on a non-self-generated voice as their primary means of communication. Synthesized voices provided by augmentative and alternative communication (AAC) devices are often unnatural-sounding, lacking in emotional prosody and slower than self-produced speech. AAC users report feeling like they present a distorted representation of themselves, which can decrease the desire to engage in social interaction and contribute to the loss of a social identity (Nathanson, 2017). Only recently has the moral case for more personalized voices been considered, with companies such as VocaliD (www.vocalid.ai) now providing a selection of voices from which AAC users might choose a voice that better conveys their self-identity.

It is important therefore to ask whether a new externally-generated voice can become associated with the self and elicit a self-bias. However, measuring the boundaries of the self has previously relied either on wholly subjective self-reports whereby participants explicitly evaluate themselves against others (see Symons & Johnston, 1997) or on comparisons of processing biases between overly familiar self-associated stimuli, such as the true self-voice, and unfamiliar, non-self-voices. Inherently, these approaches cannot address the problem of whether a new, unfamiliar self-voice can become self-associated and elicit a bias. Further, these approaches face a problematic confound of stimulus familiarity that makes it difficult to disentangle the potential bias for self-associated information from the familiarity of that information. More recent methodologies have therefore focused on creating paradigms that are both independent of stimulus familiarity and provide a more robust measure of self-bias (Cunningham, Turk, Macdonald, & Macrae, 2008; Sui, He, & Humphreys, 2012).

Sui et al. (2012) introduced a socio-associative perceptual matching task which required participants to form rapid new associations between themselves and arbitrary objects. Participants were asked to form associations between three geometric shapes (a circle, a triangle, and a square) and three identity labels (you, friend, stranger). For example, the participants were told: 'the circle represents you, the square a friend, and the triangle, a stranger'. These three shape-identity associations were thus established as 'matched' pairs. In the main task, participants were presented with pairings, each comprising one of the shapes and one of the identity labels, and asked to judge whether or not the two stimuli were correctly matched. The results consistently showed a clear perceptual bias towards the shape associated with the self, eliciting faster and more accurate responses compared to the shapes associated with either the friend or stranger (Sui et al., 2012).

Sui et al. have thus provided key insights into the nature of self-bias. Specifically, that the process of referring a stimulus to the self serves to form a rapid new self-association. Further, in becoming self-associated, the stimulus is assigned increased personal significance and becomes perceptually more salient. Neurally, this new self-association

activates increased functional connectivity between the ventral medial pre-frontal cortex, an area linked to self-representation, and the left posterior superior temporal sulcus, an area associated with social attention (Sui, Rotshtein, & Humphreys, 2013). Thus, the social salience for the self-associated stimulus is moderated and it becomes perceptually prioritized. This processing bias is termed the self-prioritization effect (Sui et al., 2012).

The ability to link new external stimuli to the self has also been explored in relation to memory. Cunningham et al. (2008; also, Cunningham, Brady-Van den Bos, & Turk, 2011) demonstrate that a stimulus can be encompassed within the self by taking ownership of it. Specifically, a collection of food items, represented pictorially, was divided into either the participant's shopping basket or another's basket. Each participant was told that they would own whichever food items were in their basket at the end. In a surprise test of memory recall, participants were more likely to remember the items they 'owned' than the items that had been placed in another's basket, despite the items being similar and the ownership being imaginary. This suggests that once an item or stimulus is associated with the self via ownership, encoding of that item is enhanced. This process may operate similarly to the self-prioritization effect demonstrated by Sui et al. (2012).

Items that we own are also perceived as having more positive traits than similar items belonging to another. Gregg, Mahadevan, and Sedikides (2017) gave participants arbitrary ownership of an abstract concept by telling them 'you have a theory about. . .' while they gave other participants an abstract concept that was 'owned' by someone else, for example, 'Alex has a theory about. . .'. When assessed, the participants who 'owned' the theory were more biased towards it; they were more likely to believe their theory was true, and also endorsed it more strongly than the participants given Alex's theory. This demonstrates that self-bias from affective-ownership can extend beyond concrete visual stimuli to abstract ideas. Although this positivity bias may reflect the consequence of encoding a stimulus to the self, rather than the mechanism, it demonstrates that the self can be extended to multiple different types of information at different levels of processing, all of which accrue a processing advantage.

Furthermore, Schäfer, Wesslein, Spence, Wentura, and Frings (2016) have demonstrated that the bias for self-associated stimuli also holds for other sensory modalities, including touch and audition. Schäfer et al. (2016) replicated the perceptual matching paradigm using an adapted version with auditory or tactile stimulus cues. In the auditory associative task, participants were asked to associate three neutrally valenced pure tones with the three identities of self, friend, and other. At test, results showed a comparable self-prioritization effect for the self-associated auditory stimuli. This study thus provides the first evidence that self-prioritization processes operate across the senses, and that we are able to incorporate a diverse range of sensory stimuli into the self.

Finally, the flexibility with which we can extend the self is further exemplified by Payne, Tsakiris, and Maister (2017), who asked participants to form new self-associations to three unfamiliar faces using the perceptual matching task. Payne et al. (2017) demonstrated that participants could successfully form an association to a new self-face, with results again showing prioritization of faces arbitrarily tagged as 'self', relative to faces tagged as belonging to friends or strangers. This result is surprising given that faces are not neutral stimuli akin to arbitrary geometric shapes. Rather, our own face is one of the most salient pieces of self-associated information we have and, importantly, unfamiliar faces should typically be processed as other-associated purely by virtue of not being our own. If multifarious pieces of self-associated information together give rise to our self-identity – an identity that is recognizably distinct from another's identity – the

incorporation of a new self-face into the representation of self seems to blur the boundary of self and other.

In demonstrating that self-bias can be attributed to an other's face, Payne et al.'s study provides the closest test case for which complex, other-related, signals can be accepted as part of the self. Voices are, like faces, highly social and identifying signals, yet they differ in that they require self-generated actions in order to exist; it is only through the varied, idiosyncratic, and deliberate use of our voice that we can share it with others to express our self-identity. Thus, to demonstrate that an unfamiliar voice generated by another person can become associated with our self and be perceptually prioritized would be a significant addition to our understanding of the self and the flexibility of our self-representation.

Here, we present three experiments using the paradigm introduced by Sui et al. (2012) to test whether the self-prioritization effect is found for unfamiliar voice identities. In Experiment 1, we establish that the effect can indeed be observed for voices as has been previously reported for other stimulus categories. In Experiments 2 and 3, we explore whether properties of the voice stimuli (i.e., matching the participant's gender identity vs not) and the method of assignment to the self (i.e., choosing the self-voice vs having no choice) would modulate the effect. These questions relating to how a voice sounds, and how it is chosen, are directly pertinent to the applied context of AAC devices; however, these are also novel questions that allow us to expand upon the existing knowledge of the self-prioritization effect for non-self-generated, social stimuli. All three experiments, including the planned analyses and predictions for each, were pre-registered via the Open Science Framework (see links below).

Experiment 1

In the first study, we asked whether a new, externally-generated voice could become associated with the self and prioritized in perception relative to voices associated with either a friend or a stranger. This experiment was pre-registered via the Open Science Framework (https://osf.io/hg3w6/).

Hypothesis

Within the perceptual matching paradigm, listeners will have quicker reaction times and greater recognition accuracy for the new self-voice compared to either the friend-voice or stranger-voice, revealing a self-prioritization effect.

General methods

The following section outlines the methods for all three studies, which utilize the same paradigm. Specifics to Experiment 1 are included thereafter.

Participants

All participants were recruited online via Prolific (www.prolific.ac) as native speakers of English with no hearing difficulties, and tested online using Gorilla (gorilla.sc, Anwyl-Irvine et al., 2019). Participants were also required to pass a headphone check (Woods,

Siegel, Traer, & McDermott, 2017) to ensure they were wearing headphones and able to hear the stimuli.

None of the participants had taken part in any of the pilot studies associated with this project and, upon completion of the study, were paid for their participation. Ethical approval was obtained from the Departmental Ethics Committee in Speech, Hearing and Phonetic Sciences at UCL (SHaPS-2018-CM-029 and SHaPS-2019-CM-030), and informed consent was obtained from all participants prior to testing.

In Experiment 1, we recruited 35 participants (mean age = 27.25 years, SD = 6.41 years, age range = 18–40, 29 male). The sample size was determined according to a previous study using the same paradigm by Schäfer et al. (2016). Their a priori power calculation suggested an effect size of dz $\geq$ .52 could be achieved given N = 31, α = .05, and a power of 1 − β = .80. A true effect size of dz = .51 was subsequently found. After exclusion criteria were applied (see 2.2.5 Pre-processing), data from 31 participants were analysed in Experiment 1.

Stimuli

Across all experiments, the stimuli included auditory exemplars of three male speakers and three female speakers each saying 'hello' in Southern Standard British English.

The male-voiced exemplars were extracted from a larger pool of original recordings, which were obtained in sound-attenuated booths using desktop computers running Audacity (https://www.audacityteam.org/) and either a Røde NT1A microphone (Røde Microphones, Sydney, Australia) or a Neumann TLM103 microphone (Neumann, Berlin, Germany). Ethical approval for these recordings was obtained (approval number 897). The female-voiced exemplars were recorded using a Rode NT-1A microphone via an RME Fireface UC audio interface. Recordings were digitized again using Audacity (2019) as 16-bit, stereo WAV files, with sample rate 44100Hz. All stimuli were normalized for RMS amplitude using PRAAT (Boersma, 2001, Boersma and Weenink, 2020).

The three male voices and three female voices were each selected from a wider pool of voices based on trait ratings obtained in pilot studies. In the first pilot study, a group of listeners (n = 15 participants, mean age = 30.33 years, SD = 6.63, age range = 18–40 years, 8 female) rated 30 auditory exemplars of 'hello' from six different male speakers for attractiveness (*'How attractive does this voice sound*?') and, separately, for trustworthiness ('How trustworthy does this voice sound?') on a 7-point Likert scale whereby 1 denoted '*not attractive/trustworthy at all'* and 7 denoted '*very attractive/trustworthy'*. We selected three male voices for use in these experiments based on their being well-matched in attractiveness (mean ratings $\pm$ SDs = 4.6 $\pm$ 1.2, 4.5 $\pm$ 1.3, 4.8 $\pm$ 1.4) and trustworthiness (mean ratings $\pm$ SDs = 4.6 $\pm$ 1.2, 5.2 $\pm$ 1.1, 4.7 $\pm$ 1.2).

In the same pilot, participants were asked to rank the six different male voices in order of preference, 1 to 6, whereby 1 denoted '*The voice I would most like to have as my own*' down to 6, which denoted '*The voice I would least like to have as my own*'. A Friedman Test was then run in the R environment using *stats* (R Core Team, 2013) to determine whether the six voices differed significantly in their mean ranking. The results indicated that the voices did differ significantly ($\chi^2(5)$ = 23.82, p < .001) so further Nemenyi post-hoc tests were run using *PMCMR* (Pohlert, 2014). The comparisons revealed that, of the original six voices, the three male voices used across these three experiments were well-matched in their mean ranking with all p-values >.05.

In a second pilot study, six female voices were also trait rated and ranked for preference in the same way (n = 15 participants, mean age = 27.6 years, SD = 6.31, age

range = 18–39 years, 6 female). As with the male voice stimuli, the three female voices selected for use in Experiments 2 and 3 were chosen for being well-matched for attractiveness (mean ratings $\pm$ *SD*s = 4.4 $\pm$ 1.0, 4.2 $\pm$ 1.4, 4.8 $\pm$ 1.4) and for trustworthiness (mean ratings $\pm$ *SD*s = 4.8 $\pm$ 1.2, 4.6 $\pm$ 1.3, 4.9 $\pm$ 1.0).

Finally, the same participants were asked to rank the six different female voices in order of preference. The ranking data were analysed with a Friedman Test in the *R* environment, and the results indicated that the voices differed significantly ($\chi^2(5) = 28.83, p < .001$). Further Nemenyi post-hoc tests were run and the comparisons revealed that, of the original six voices, the three female voices used in Experiments 2 and 3 were well-matched in their mean ranking with all *p*-values > .05.

Experiment 1 used just the male voice stimuli, with two auditory tokens of 'hello' per voice.

Procedure

The paradigm involved two stages: the familiarization phase and the test phase. The online tasks here described can be accessed and previewed online via Gorilla Open Materials at: https://gorilla.sc/openmaterials/45935.

Familiarization stage.   In the familiarization stage, participants were passively exposed to each of the three voices and their associated identity label. Following a 500 ms fixation cross, the word 'you', 'friend', or 'stranger' was displayed in the centre of the screen in black uppercase font on a white background and remained on-screen for 3000 ms. 500 ms after the label's onset, an auditory exemplar of 'hello' from the to-be-associated voice was played, lasting approximately 500 ms. After the auditory stimulus finished, the label remained on the screen until the end of the 3000 ms trial. The familiarization phase consisted of 12 trials, with each label-voice pairing being presented four times, and each of the two auditory exemplars per voice played twice. Stimuli were presented in a random order. The familiarization stage thus lasted approximately 1 minute, including on-screen instructions, and was immediately followed by the test phase.

Test phase.   The test phase started with 12 practice trials. Each trial started with a 500 ms fixation cross in the centre of the screen, after which an auditory exemplar played for its total duration (approx. 500 ms). Immediately after the auditory offset, the word 'you', 'friend', or 'stranger' was displayed in the centre of the screen. Participants were asked to judge whether the identity label on-screen was a match or mismatch to the voice heard and to respond by keyboard press as quickly and as accurately as possible. Participants were instructed to press the left arrow for 'MATCH' and the right arrow for 'MISMATCH'. This left-right ordering of match and mismatch remained constant on all trials, for all participants. In the practice trials only, text appeared on-screen to remind participants that a left button press meant a match, and a right button press meant mismatch. In all test trials, however, no visual reminder was provided. Feedback was given on-screen for 500 ms immediately following every response: a green tick for correct, a red cross for incorrect, and text feedback of 'TOO SLOW' for responses occurring after 1500 ms. The next trial began after the 500 ms feedback period was complete.

Following the practice trials, participants performed three blocks of 72 trials (216 trials in total) and were informed of their overall accuracy at the end of each block. The order of

match vs mismatch trials was randomized for all participants. The experiment lasted approximately 25 minutes in total and participants were debriefed at the end.

Design
Experiment 1 consisted of a within-subjects factor, 'voice identity' (self vs friend vs stranger voice) × a within-subjects factor, 'trial type' (match vs mismatch) design.

Pre-processing
In each experiment, we analysed three measures: reaction times, sensitivity, and accuracy. The first two measures were included to align with previous reports of the self-prioritization effect by Sui et al. (2012) and Schäfer et al. (2016). We also included trialwise accuracy to allow us to run a complementary analysis to the RT mixed model that included the same random effects structure.

Reaction times: RTs were measured as the delay between the onset of the visual label and the participant's response and categorized by the voice identity (self-voice, friend-voice, stranger-voice) per trial, and not by the category of the presented label ('YOU', 'FRIEND', 'STRANGER'). All erroneous responses, as well as responses shorter than 200ms or longer than 1500ms were excluded from the RT analysis in line with Sui et al.'s (2012) approach.

Sensitivity: unbiased d' scores were calculated by combining performance scores from match and mismatch trials at each level of the voice identity factor (self, friend, stranger). Specifically, YES/NO responses were recoded into hits and false alarms, following the log-linear approach to adjust for cases involving 100% hits or 0% false alarms (Stanislaw & Todorov, 1999).

Accuracy: Trialwise accuracy scores were recorded for each participant. Whole data sets for participants showing overall performance accuracy at chance ($\leq 50\% + 95\%$ CI) were excluded as this indicated random responses.

In Experiment 1, 2.48% of the trials were eliminated on the RT criteria above as well as whole data sets for two participants with more than 50% of responses shorter than 200 ms. A further two data sets were eliminated on the accuracy criterion. None of these participants was replaced; therefore, data from a total of 31 participants were analysed in Experiment 1.

Analysis
Across all experiments, reaction times and sensitivity were assessed with linear mixed models (LMM) using *lme4* (Bates, Maechler, Bolker, & Walker, 2014) in the *R* environment (R Core Team, 2013). Accuracy was assessed using binomial generalized linear mixed models (GLMM). In every analysis (with the exception of sensitivity in Experiment 1), we ran a model that included an interaction between voice identity and the manipulated variable, two fixed effects (voice identity, and the manipulated variable: trial type in Experiment 1, gender-matching in Experiment 2, and choice in Experiment 3), and a random intercept of 'participant'. Statistical significance of the effects was established via likelihood ratio tests by dropping effects of interest from the appropriate model. For example, to establish whether the interaction was significant, we dropped this interaction from a model including both fixed effects. To test for the significance of either of the main effects, we dropped the relevant effect from a model that only included the two main

effects. For all analyses reported here, post-hoc comparisons were conducted in *emmeans* (Lenth, 2016) and were adjusted for the multiple comparisons via Bonferroni correction.

For GLMMs, odds ratios and associated confidence intervals are provided as an estimate of the size of the relevant effects. An odds ratio of 1 indicates that no effect is present, while odds ratios that deviate from 1 indicate an effect is present. The larger the deviation, the bigger the effect. If the confidence intervals do not cross 1, the relevant effect is significant. For LMMs, the models' estimates and associated confidence intervals for each effect are reported. The further away from 0 the estimates are, the bigger the effect. If the confidence intervals do not cross 0, the relevant effect is significant.

In addition to the linear mixed models, we also ran ANOVAs on all three measures to enable a more direct comparison with previous published studies. The results can be found in Supplemental Materials.

Results & discussion

Descriptive statistics for reaction times, accuracy, and sensitivity scores (d') for the different conditions and voice identities are given in Table 1.

Reaction times

The reaction time data are plotted in Figure 1. We ran an LMM on RTs to assess whether the self-voice had been perceptually prioritized relative to the others depending on the trial type (match trials vs mismatch trials). For the post-hoc tests, alpha was set as .008 after Bonferroni correction for six comparisons.

lmer (reaction time $\sim$ voice identity * trial type + (1 | participant), REML = $'$FALSE$'$)

Full model outputs are reproduced in Table S1. There was a significant interaction between voice identity and trial type ($\chi^2(2) = 21.32, p < .001$). Post-hoc tests confirmed that reaction times for the self-voice were significantly quicker than those for the friend-voice ($p < .001$) and the stranger-voice ($p < .001$). However, this was only the case within matched trials, which is in line with previous studies. There was no significant difference between reaction times for the friend-voice and stranger-voice in either match,

Table 1. Mean RTs, accuracy, and d-primes for Experiment 1

Condition	Voice Identity	Mean RT (ms)	Accuracy	d' [a]
Match trials	Self	584 (211)	0.91 (0.29)	2.64 (1.11)
	Friend	635 (220)	0.88 (0.33)	2.31 (0.97)
	Other	645 (227)	0.87 (0.34)	2.29 (1.06)
Mismatch trials	Self	714 (242)	0.83 (0.38)	
	Friend	722 (217)	0.81 (0.39)	
	Other	712 (216)	0.81 (0.40)	

Note. RT = reaction time; Accuracy = proportion correct.
Standard deviations appear within parentheses.
[a]Performance scores from match and mismatch trials are combined to provide d' scores.
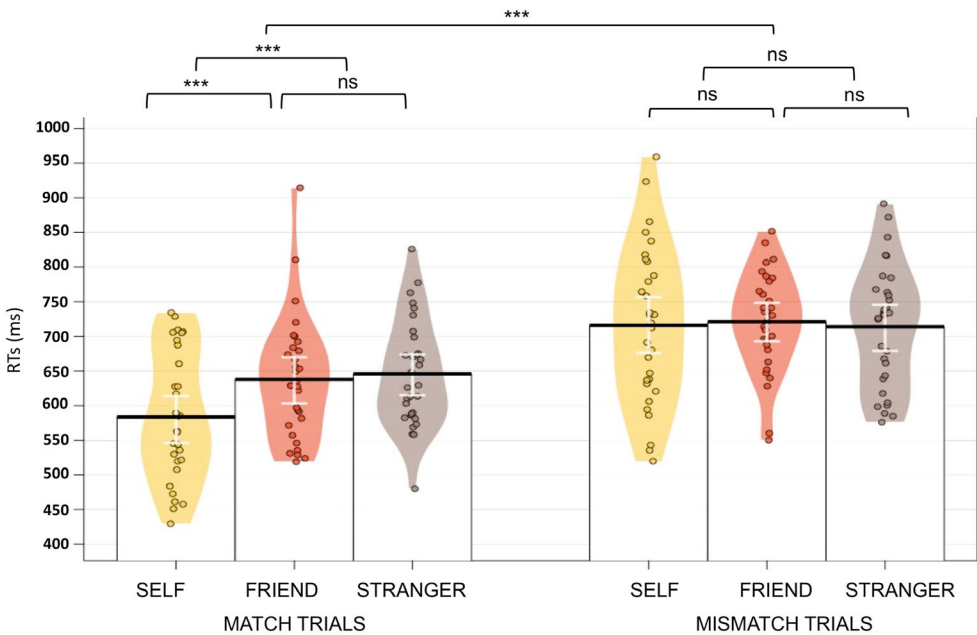
Figure 1. Experiment 1: Mean reaction times (RT) as a function of the voice identity (self, friend, or stranger) and the trial type (match vs mismatch). The error bars indicate the SEs of the means. Coloured segments show smoothed density curves for the full data distribution, while individual dots indicate mean RTs per participant. Horizontal bars show post-hoc comparisons with asterisks denoting significance within RTs as determined via likelihood ratio tests (see Results). ***p < .001. [Colour figure can be viewed at wileyonlinelibrary.com]

or mismatch trials (*p*s > .008). These results show therefore a significant self-prioritization effect for the perception of voices, thus supporting our hypothesis.

Sensitivity
We further assessed the self-prioritization effect by running an LMM on d' scores to assess performance using the model below. For the post-hoc tests, alpha was set as .017 after Bonferroni correction for three post-hoc comparisons.

$$\text{lmer (sensitivity} \sim \text{voice identity} + (1|\text{participant}), \text{REML}='\text{FALSE}')$$

Full model outputs are reproduced in Table S2. There was again a significant main effect of voice identity ($\chi^2(2) = 6.24, p = .044$). However, post-hoc comparisons showed that although sensitivity for the self-voice was numerically greater than that for both the friend-voice and stranger-voice, the difference did not reach significance (*p*s > .017) after Bonferroni correction.

Accuracy
We also ran a GLMM on trialwise accuracy. For the post-hoc tests, alpha was set as .017 after Bonferroni correction for three post-hoc comparisons.

glmer (accuracy $\sim$ voice identity * trial type $+$ (1 | participant), family $= '$binomial$'$)

Full model outputs are reproduced in Table S3. The interaction of voice identity and trial type was not significant ($\chi^2(2) = 2.29$, $p = .318$). There was, however, a significant main effect of voice identity ($\chi^2(2) = 11.84$, $p = .003$), showing that performance accuracy for the self-voice was higher than for both the friend-voice ($p = .013$) and the stranger-voice ($p < .001$), while the friend-voice and stranger-voice did not significantly differ. There was also a main effect of trial type ($\chi^2(1) = 70.63$, $p < .001$) showing that accuracy was significantly higher in match trials than in mismatch trials ($p < .001$); a result which falls in line with previous studies.

In Experiment 1, we asked whether an unfamiliar voice could become self-associated and we measured the processing bias subsequently afforded to this new self-voice relative to voices newly associated with a friend or stranger. Our results showed a clear self-prioritization effect, with quicker reaction times, higher accuracy, and increased perceptual sensitivity to the self-voice relative to other voice identities. This therefore supports our hypothesis and provides a pattern of results that mirrors those reported by Sui et al. (2012), Schafer et al. (2016), and Payne et al. (2017). This demonstrates for the first time that our 'self' can quickly be extended to incorporate an unfamiliar voice – that inherently belongs to another – as our own. We further show that voices do not need to be self-generated in order to be self-associated and perceptually prioritized as a self-relevant signal. This has positive implications for the potential uptake of synthesized voices by patients with vocal impairments.

Experiment 2

Having demonstrated that there is a self-prioritization effect in the perception of voices, Experiment 2 asked whether this effect is increased if the new self-voice is gender-matched to the identity of the listener. This experiment was pre-registered via the Open Science Framework (https://osf.io/r6g2m/) and can be previewed via Gorilla Open Materials at: https://gorilla.sc/openmaterials/46084.

People prefer others whom they perceive or imagine to be similar to themselves because it activates the positive and implicit associations they already hold about themselves (Jones, Pelham, Carvallo, & Mirenberg, 2004). This perceived similarity reduces the social distance between them and, relatedly, the degree to which the self and the other are differentially processed (Benoit, Gilbert, Volle, & Burgess, 2010). For instance, people are able to recall objects belonging to a similar other better than objects belonging to a dissimilar other (Allan, Morson, Dixon, Martin, & Cunningham, 2017). This shows that the perceived closeness of an object's association to the self can influence the way we process it: more closely associated items are more prioritized in memory.

It is reasonable therefore to suggest that an other-generated voice that is more similar to the true self-voice of the listener – in terms of gender-related acoustics – may be perceived as more closely related to the self than a voice that represents a less similar other. Voice characteristics arise from, and subsequently convey, information about our physical self-identity including body size, shape, health and gender. In adult humans, sexual dimorphism in voice characteristics – including fundamental frequency and formant spacing – is extremely large in comparison with other species (Puts et al., 2016). Further, the differences in these voice characteristics are perceptually highly salient (Lass, Hughes, Bowyer, Waters, & Bourne, 1976; Owren, Berkowitz, & Bachorowski, 2007).

Thus, through manipulating the sex of the experimental voice identities, and their match to the gender identity of participants, we can effectively test whether an externally-generated vocal signal that better fits within the parameters of the participant's vocal apparatus and self-identity may be deemed more personally relevant than a voice that signals a different gender identity.

Hypothesis
We hypothesize that the self-prioritization effect for the self-voice will increase when it is matched to the gender identity of the listener. Such an increase should result in quicker reaction times and higher accuracy in response to the self-voice, relative to the friend- or stranger-voice, when the self-voice is matched to the gender identity of the listener as compared to when the self-voice is not matched to the gender identity of the listener.

Methods
Unless stated below, the methods for Experiment 2 replicated methods from Experiment 1.

Participants
96 participants took part (48 females: mean age = 29.1 years, *SD* = 6.28 years, age range = 18–40; 48 males: mean age = 27.4 years, *SD* = 6.58 years, age range = 18–40). The data from participants (*N* = 2) showing overall performance accuracy at chance (≤50% + 95% CI) were excluded and replacement participants sampled such that data from 96 participants were included in the reported analyses.

Randomization
Male and female participants were randomly allocated to either a 'gender-matched' group (*N* = 48: 24M, 24F) or a 'non-gender-matched' group (*N* = 48: 24M, 24F), whereby the gender of the participant would match or not match the gender of the voices.

Stimuli
In addition to the male voice stimuli used in Experiment 1, we included female voice stimuli from three SSBE-accented speakers.

Design
The study consisted of a within-subjects factor, 'voice identity' (self vs friend vs stranger) × a between-subjects factor, 'gender-matching' (gender-matched vs non-gender-matched) design. Further, a within-subjects factor of trial type was present (match vs mismatch trials).

Pre-processing
Based on the exclusion criteria reported earlier, fewer than 1% of the trials were eliminated.

Table 2. Mean RTs, accuracy and d-prime scores in matched trials for Experiment 2

Condition	Voice Identity	Mean RT (ms)	Accuracy	d' [a]
Gender-matched	Self	541 (203)	0.92 (0.27)	2.89 (0.98)
	Friend	572 (213)	0.92 (0.27)	2.90 (1.01)
	Other	592 (221)	0.85 (0.35)	2.49 (0.97)
Non-gender-matched	Self	536 (205)	0.94 (0.24)	2.95 (1.04)
	Friend	570 (222)	0.91 (0.29)	2.87 (1.16)
	Other	589 (224)	0.88 (0.33)	2.68 (1.21)

Note. RT = reaction time; Accuracy = proportion correct.
Standard deviations appear within parentheses.
[a]Performance scores from match and mismatch trials are combined to provide d' scores.

Analysis
In Experiment 1, we found that the self-prioritization effect emerged on match trials only (i.e., trials in which the presented voice and identity label are correctly 'matched' according to their learnt association). Therefore, following the logic of Sui et al. (2012) and Schafer et al. (2016), the analyses of reaction times and raw accuracy are henceforth reported separately for match and mismatch trials.

Results & discussion
Descriptive statistics for reaction times, accuracy, and sensitivity scores (d') for the different conditions and voice identities are given in Table 2.

Reaction times
The reaction time data are plotted in Figure 2. As with analysis in Experiment 1, we ran an LMM on RTs to assess whether the self-voice had been perceptually prioritized. First, we analysed match trials only. For the post-hoc tests, alpha was set as .017 after Bonferroni correction for three post-hoc comparisons.

$$\text{lmer (Reaction time} \sim \text{voice identity} * \text{gender} - \text{matching}$$
$$+ (1\,|\,\text{participant)}, \text{REML} = '\text{FALSE}')$$

Full model outputs are reproduced in Table S4. The interaction of voice identity and gender-matching was non-significant ($\chi^2(2) = .27, p = .875$). There was, however, again a main effect of voice identity ($\chi^2(2) = 106.78, p < .001$): In both groups, reaction times for the self-voice were significantly quicker than those for either the friend-voice ($p < .001$) or the stranger-voice ($p < .001$), and reaction times to the friend voice were also significantly quicker than those for the stranger-voice ($p < .001$). Lastly, we found no main effect of gender-matching on reactions times ($\chi^2(1) < .01, p = .934$).[1]

These RT data therefore reveal a significant self-prioritization effect in the perception of voices, but show that prioritization was not modulated according to whether the self-voice was matched to the gender identity of the listener.

---

[1] *The effect of listener gender was also analysed for Experiments 2 and 3 and was found to be non-significant in both cases.*
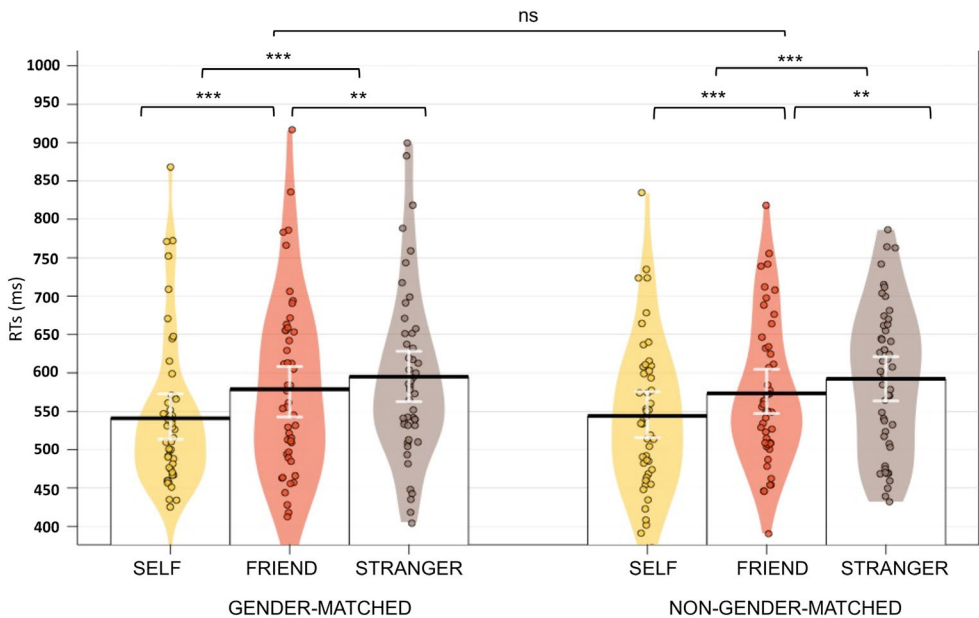
Figure 2. Experiment 2: Mean reaction times (RT) as a function of the voice identity (self, friend, or stranger) and gender-matching (matched vs non-matched, i.e., whether the voice matches the gender identity of the listener, or not). The effect of gender-matching was not significant. Graph models match trials only. The error bars indicate the SEs of the means. Coloured segments show smoothed density curves for the full data distribution, while individual dots indicate mean RTs per participant. Horizontal bars show post-hoc comparisons with asterisks denoting significance within RTs as determined via likelihood ratio tests (see Results). ***$p < .001$. [Colour figure can be viewed at wileyonlinelibrary.com]

Identical analyses of RT data were then run on the mismatch trials, but the interaction effect ($\chi^2(2) = 1.03$, $p = .596$), the effect of voice identity, ($\chi^2(2) = .07$, $p = .968$), and the effect of gender-matching ($\chi^2(2) = .69$, $p = .708$) were all non-significant.

Sensitivity
We also ran an LMM on d' scores. For the post-hoc tests, alpha was set as .017 after Bonferroni correction for three post-hoc comparisons.

$$\text{lmer}\,(\text{sensitivity} \sim \text{voice identity} * \text{gender} - \text{matching}$$
$$+ (1\,|\,\text{participant}),\ \text{REML} = {'}\text{FALSE}{'})$$

Full model outputs are reproduced in Table S5. The interaction of voice identity and gender-matching was non-significant ($\chi^2(2) = 1.57$, p = .455). There was again a main effect of voice identity ($\chi^2(2) = 16.62$, $p < .001$), with post-hoc comparisons revealing that, across groups, d-primes for the self-voice were significantly higher than those for the stranger-voice ($p < .001$) but did not differ significantly from the friend-voice. Further, d-primes for the friend-voice were significantly higher than those for stranger-voice ($p < .001$). There was again no main effect of gender-matching on reaction times ($\chi^2(1) = .16$, $p = .687$).

Accuracy

Finally, we ran a GLMM on trialwise accuracy. Here we analysed the match trials. For the post-hoc tests, alpha was set as .008 after Bonferroni correction for six comparisons.

$$\text{glmer (accuracy} \sim \text{voice identity} * \text{gender} - \text{matching}$$
$$+ (1 \,|\, \text{participant}), \text{family} = \,'\text{binomial}')$$

Full model outputs are reproduced in Table S6. There was a significant interaction between voice identity and gender-matching on accuracy ($\chi^2(2) = 7.09, p = .029$). Post-hoc comparisons revealed that accuracy for the self-voice was significantly higher than accuracy for the stranger-voice ($p < .001$) in both gender-matched and non-gender-matched groups. However, accuracy for the self-voice only differed from the friend-voice in the non-gender-matched group ($p = .001$). Accuracy for the self-voice did not differ significantly when it was gender-matched in comparison to when it was not.

Identical analyses within the mismatch trials revealed that neither the interaction effect ($\chi^2(2) = 1.03, p = .596$), the fixed effect of voice identity ($\chi^2(2) = .065, p = .968$), nor the fixed effect of gender-matching ($\chi^2(1) = .069, p = .793$) were significant.

In Experiment 2, we asked whether the self-prioritization effect could be modulated by the self-voice being gender-matched to the identity of the listener. Our results show that gender-matching did not significantly influence prioritization here, with equivalent reaction times and perceptual sensitivity to the voices when they were matched as to when they were mismatched. Although we observed a significant interaction of voice identity and gender-matching for accuracy scores, this did not manifest in interpretable differences in terms of how the self-voice related to the friend and stranger voices across the two groups. Overall, these results suggest that the perceived physiological similarity between the novel voices and the participant's own voice had no effect on the self-prioritization effect. This is unexpected (see the General Discussion for suggested explanations). Instead, prioritization was similar for both groups, in that perceptual processing was biased towards the new self-voice relative to the friend- or stranger-voice, thereby supporting the results from Experiment 1.

Experiment 3

Experiment 3 further investigated which factors may influence the acceptability of an externally-generated voice being integrated into self-representation. Specifically, we asked whether the action of choosing a new voice for the self increases the level of bias towards it, and the implications of this on the prioritization effect. This experiment was pre-registered via the Open Science Framework (https://osf.io/wz2a m/) and can be previewed via Gorilla Open Materials at https://gorilla.sc/openmateria ls/46086.

The effect of choice on self-bias has been investigated in the memory domain by Cloutier and Macrae (2008). They demonstrate that participants who are given a choice over which items to own can better recall those items relative to participants who had no choice and were instead afforded ownership by someone else. The act of selection, of making a personal choice over external stimuli, is associated with enhanced memory for the later outcomes of those choices (Cloutier & Macrae, 2008). Further, self-selected items are more likely to align with the current preferences and biases of the self, and so be more strongly encoded to the self – and more prioritized – than unchosen items.

   The extent to which the amount of personal choice we have influences our degree of bias has also been explored by Cunningham et al. (2008) and Cunningham et al. (2011). They demonstrate that bias towards an item only occurs when it is both 'self-owned' and 'self-chosen', showing that having more personal choice increases bias. This is mirrored in Huang, Wang, and Shi's (2009) study, in which the bias that was afforded by merely self-owning an item (relative to an other-owned item) disappeared when participants were given a chance to compare their own items with another's. In contrast, the bias afforded by self-*choosing* an item, either to become a self-owned item or to give as an other-owned item, remained stable regardless of the ownership. Thus, encoding for a self-chosen item seems to be stronger and able to override ownership biases.

Hypothesis

Experiment 3 therefore predicts that the self-prioritization effect in voices will be increased if the new self-voice is first chosen by the listener, relative to when it has been randomly allocated. Such an increase in prioritization should result in quicker reaction times and higher accuracy in response to the self-voice – in comparison with the voices associated with the friend or stranger – when the new self-voice has been chosen, relative to when it has been randomly allocated.

Methods

Unless stated below, the methods for Experiment 3 replicated methods from Experiment 1 and 2.

Participants

96 new participants were recruited, forming a 'choice' group (48 females: mean age = 29.2 years, SD = 6.27 years, age range = 18–40; 48 males: mean age = 26.5 years, SD = 6.71 years, age range = 18–40). As in previous analysis, data sets for participants showing overall performance accuracy at chance ($\leq$50% + 95% CI) were excluded (N = 3), as this indicated random responses. These participants were replaced to reach the full sample size of 96 participants. Data from these 96 participants were then combined with data from the 96 participants from Experiment 2 who had been randomly allocated voices, thus forming the 'no choice' group (48 females: mean age = 29.1 years, SD = 6.28 years, age range = 18–40; 48 males: mean age = 27.4 years, SD = 6.58 years, age range = 18–40). Overall, data from 192 participants were included in this study.

Stimuli

Experiment 3 used both male and female voice stimuli, as in Experiment 2.

Design

The study consisted of a factorial design with one within-subjects factor, 'voice identity' (self vs friend vs stranger) and one between-subjects factor, 'choice' (choice vs no choice). Further, a within-subjects factor of trial type was present (match vs mismatch trials).

Procedure

For the 'choice' group, Experiment 3 began with a voice selection task in which the participants were presented with two tokens of 'hello' from each of the three male voices or three female voices, labelled on-screen as 'Voice A', 'Voice B', and 'Voice C'. They were then asked to choose which voice they wanted to represent them for the rest of the study. The voices for 'friend' and 'stranger' were allocated randomly, and evenly, thereafter. Second, participants were asked to rate the strength of their preference for their chosen voice on a 5-point scale where 1 denoted, '*Not strong at all: I was indifferent/picked randomly*' and 5 denoted, '*Very strong: I preferred my chosen voice much more than the others.*'

Thereafter, the procedure replicated that of Experiments 1 and 2.

Pre-processing

In the self-voice selection task, participants were presented with a choice of three male voices or three female voices. For participants choosing from male voices, there was a relatively balanced selection across the three identities. Voice 1 was chosen by 37.5% of participants (8F, 10M), Voice 2 was also chosen by 37.5% of participants (8F, 10M), and Voice 3 was chosen by 25% of participants (8F, 4M). The strength of preference task showed that 87.5% of participants stated they had a preference strength of at least three (out of five) for the voice they chose, with only 2.08% stating that they were indifferent to their self-voice.

Uptake of the female voices as the new self-voice was less balanced; Voice 1 was chosen by 62.5% of participants (16F, 15M), Voice 2 was chosen by 12.5% (4F, 2M), and Voice 3 was chosen by 25% of participants (4F, 7M). The strength of preference task showed that 95.8% of participants stated they had a preference strength of at least three (out of five) for the voice they chose.

Since we posit that encoding of a stimulus to the self is strengthened through making a personal choice – rather than by the strength of preference for that stimulus – data from all participants were analysed regardless of their stated preference strength.

1.18% of the trials were eliminated from the RT analysis based on our exclusion criteria.

Analysis

For the analyses of reaction times and raw accuracy, results within match and mismatch trials are reported separately, following the logic of Sui et al. (2012) and Schafer et al. (2016). Matched trials are those in which the presented voice and identity label are correctly 'matched' according to their learnt association.

Results & discussion

Descriptive statistics for reaction times, accuracy, and sensitivity scores (d') for the different conditions and voice identities are given in Table 3.

Reaction times

The reaction time data are plotted in Figure 3 for the match trials and in Figure 4 for the mismatch trials. We ran an LMM on RT data to assess whether the self-voice had been perceptually prioritized relative to the others depending on choice (whether the self-voice was chosen or not). First, we analysed match trials only. For the post-hoc tests, alpha was set as .008 after Bonferroni correction for six comparisons.

Table 3. Mean RTs, accuracy, and d-prime scores in matched trials for Experiment 3

Condition	Voice Identity	Mean RT (ms)	Accuracy	d' [a]
Choice	Self	542 (204)	0.94 (0.25)	3.10 (1.06)
	Friend	610 (231)	0.90 (0.30)	2.83 (1.15)
	Other	606 (234)	0.87 (0.33)	2.66 (1.19)
No Choice	Self	539 (204)	0.93 (0.27)	2.92 (1.01)
	Friend	571 (218)	0.92 (0.28)	2.89 (1.09)
	Other	590 (222)	086 (0.34)	2.59 (1.10)

Note. RT = reaction time; Accuracy = proportion correct.
Standard deviations appear within parenthesis.
[a]Performance scores from match and mismatch trials are combined to provide d' scores.

$$\text{lmer (reaction time} \sim \text{voice identity * choice} + (1\,|\,\text{participant}), \text{REML} = {'FALSE'})$$

Full model outputs are reproduced in Table S7. There was a significant interaction between voice identity and choice ($\chi^2(2) = 27.56$, ($p < .001$), showing reaction times were quicker for the self-voice than for friend ($p < .001$), or stranger ($p < .001$) in both groups. We hypothesized that an interaction would arise because of a difference in RTs to the self-voice when it is chosen in comparison to when it is not. However, instead, the interaction arose from a difference in RTs to the friend-voice when the self-voice was
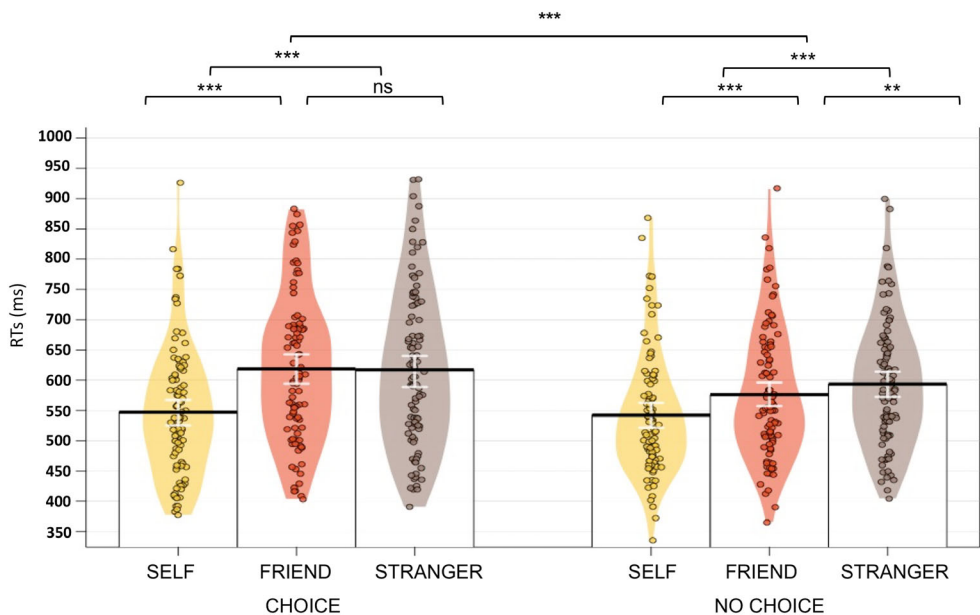
Figure 3. Experiment 3: Mean reaction times (RT) as a function of the voice identity (self, friend, or stranger) and choice (choice vs no choice, i.e., whether the self-voice is self-selected or randomly allocated). Graph models match trials only. The error bars indicate the SEs of the means. Coloured segments show smoothed density curves for the full data distribution, while individual dots indicate mean RTs per participant. Horizontal bars show post-hoc comparisons with asterisks denoting significance within RTs as determined via likelihood ratio tests (see Results). ***$p < .001$. [Colour figure can be viewed at wileyonlinelibrary.com]
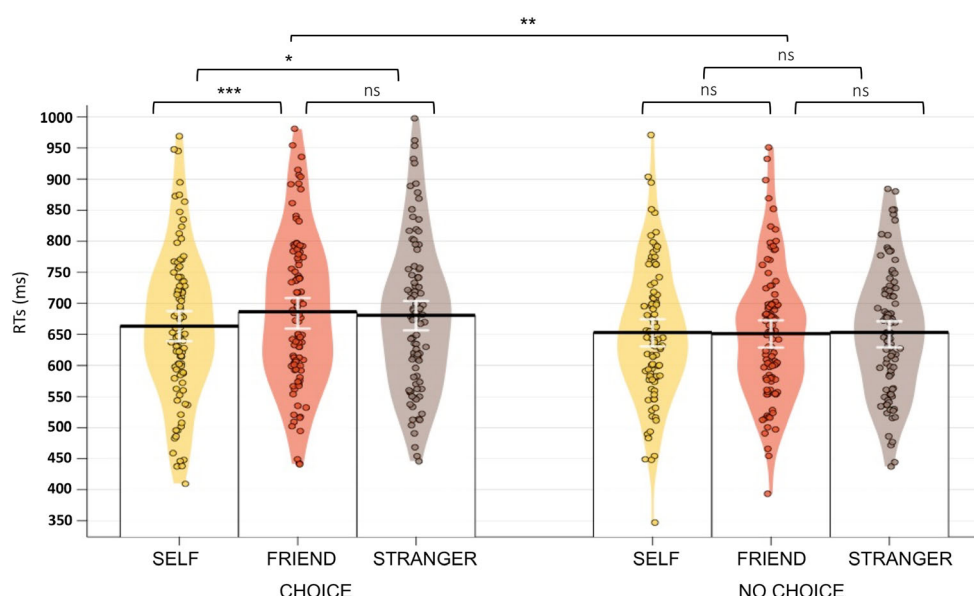
Figure 4. Experiment 3: Mean reaction times (RT) as a function of the voice identity (self, friend, or stranger) and choice (choice vs no choice, i.e., whether the self-voice is self-selected or randomly allocated) in mismatch trials only. The error bars indicate the SEs of the means. Coloured segments show smoothed density curves for the full data distribution, while individual dots indicate mean RTs per participant. Horizontal bars show post-hoc comparisons with asterisks denoting significance within RTs as determined via likelihood ratio tests (see Results). ***$p < .001$. [Colour figure can be viewed at wile yonlinelibrary.com]

chosen in comparison to when it was not. Specifically, when the self-choice is chosen the RTs for the friend voice increase such that they become equivalent to those of the stranger's voice ($p = .577$). This leaves the self-voice relatively more prioritized than either of the other two voices, which is not the case when the self-voice is randomly allocated. When the self-voice is randomly allocated, reaction times to friend and stranger ($p < .001$) are significantly different such that there is less prioritization of the self-voice relative to the other voices.

Identical analyses were run on RTs in mismatch trials. For the post-hoc tests, alpha was set as .008 after Bonferroni correction for six comparisons. These analyses also showed a significant interaction between voice identity and choice ($\chi^2(2) = 10.45$, $p = .005$). Reaction times were significantly quicker for the self-voice than for either the friend-voice ($p < .001$) or the stranger-voice ($p = .001$), but this self-prioritization effect was only present when the self-voice was chosen. When the self-voice was randomly allocated, performance did not significantly differ for the self, friend, and stranger voices. This result supports the notion that choosing the self-voice does increase the prioritization of it, thus making it more prominent relative to either of the other voices.

Sensitivity

We ran an LMM on d' scores. For the post-hoc tests, alpha was set as .017 after Bonferroni correction for three post-hoc comparisons.

$$\text{lmer (sensitivity} \sim \text{voice identity} * \text{choice} + (1 \,|\, \text{participant)}, \text{REML} = \text{'FALSE'})$$

Full model outputs are reproduced in Table S8. The interaction of voice identity and choice was non-significant ($\chi^2(2) = 3.55, p = .169$). There was, however, a main effect of voice identity on reaction times ($\chi^2(2) = 34.55, p < .001$), revealing that d-primes for the self-voice were significantly higher than those for the stranger-voice ($p < .001$) but did not differ significantly from the friend-voice ($p = .020$). D-prime scores also differed significantly between the friend-voice and stranger-voice ($p < .001$). Lastly, we tested for the significance of choice as a fixed effect; however, this was non-significant ($\chi^2(1) = .26, p = .612$).

Accuracy

Finally, we ran GLMMs on trialwise accuracy, first in the match trials. For the post-hoc tests, alpha was set as .008 after Bonferroni correction for six comparisons

$$\text{Full model : glmer (accuracy} \sim \text{voice identity} * \text{choice}$$
$$+ (1 \,|\, \text{participant)}, \text{family} = \text{'binomial'})$$

Full model outputs are reproduced in Table S9. There was an interaction between choice and voice identity ($\chi2(2) = 6.56, p = .038$). Pairwise comparisons reveal that when the self-voice was chosen, accuracy for the self-voice becomes significantly greater than accuracy for either the friend-voice ($p < .001$) or the stranger-voice ($p < .001$) showing a clear self-prioritization effect. However, when the self-voice is randomly allocated, accuracy for the self-voice does not differ significantly from the friend-voice ($p = .020$) though does differ from the stranger-voice ($p < .001$). Thus, the interaction effect arises from a relative deprioritization of the friend-voice when the self-voice is chosen compared to when it is not. This pattern of results mirrors the interaction effect found in the RTs: here too, performance for the friend-voice is reduced in the choice condition, leaving the chosen self-voice relatively more prioritized than either of the other two voices. This is not the case when the self-voice is randomly allocated. Lastly, across both groups, accuracy for the friend-voice was significantly higher than accuracy for the stranger-voice (ps < .001).

Identical analyses were also run on accuracy in mismatch trials. For the post-hoc tests, alpha was set as .008 after Bonferroni correction for six comparisons. There was a significant interaction effect here too ($\chi^2(2) = 7.98, p = .018$). The interaction effect arose from significantly higher accuracy for the self-voice than for the stranger-voice ($p = .005$) only when the self-voice was chosen. When the self-voice was randomly allocated, performance did not significantly differ for the self, friend, and stranger voices. This result supports the notion that choosing the self-voice does increase the prioritization of it, thus making it more prominent relative to either of the other voices.

Exploratory analyses

As a means to further explore the implication of choice on self-prioritization, we indexed the degree of bias for the self-voice (following Sui & Humphreys, 2017) in both the choice and no choice conditions. A larger score indicates a larger bias, whereas smaller scores indicate a relatively smaller bias. For each participant, a self-bias score can be calculated by

taking the difference in RTs for the self-voice versus the friend-voice, dividing it by the sum of RTs across the two conditions, and finally multiplying the result by 100 to achieve a percentage score. A friend-bias score can also be calculated by taking the difference in RT performance for the friend-voice relative to the stranger-voice, dividing it by the sum of the RTs across the conditions and multiplying it by 100. Bias scores are plotted in Figure 5.

The bias scores show that there is a larger bias – and greater prioritization – for 'self' relative to 'friend' across both conditions. Importantly though, when the self-voice has been chosen, self-prioritization increases and the prioritization of friend decreases relative to when the self-voice has not been chosen. This supports the earlier findings that there is a greater prominence of self, relative to others, when the stimulus is personally selected. In order to test the significance of this, we calculated the difference between the self-bias and friend-bias score per participant and then ran an independent samples t-test on the resultant scores, comparing the choice vs no choice groups. This test was significant (t (190) = −3.05, p = .003).

In Experiment 3, we asked whether the self-prioritization effect could be modulated by choice. Our results show that choice does interact with prioritization: performance for the friend-voice is significantly reduced as measured by RTs and accuracy – such that it is more distinguished from the self-voice and equivalent to those of the stranger's voice – only when the self-voice is chosen. The results suggest that when the self-voice is chosen,
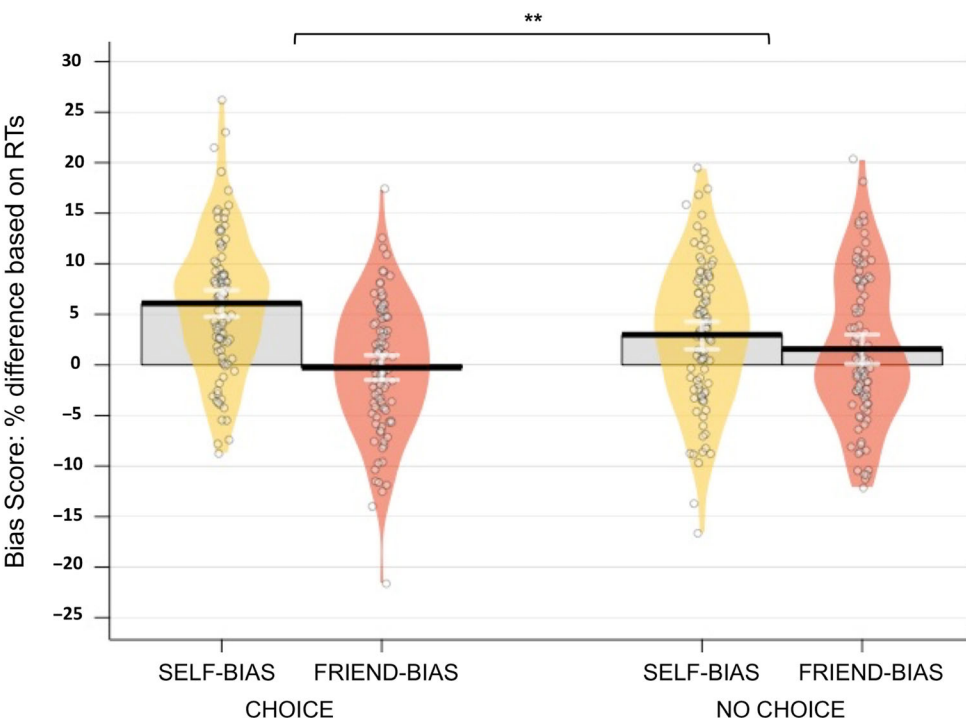

Figure 5. Exploratory analyses, Experiment 3: Measures of bias (self-bias vs friend-bias) as a function of choice (choice vs no choice, i.e., whether the self-voice is self-selected or randomly allocated). The error bars indicate the SEs of the means. Coloured segments show smoothed density curves for the full data distribution, while individual dots indicate mean scores per participant. [Colour figure can be viewed at wileyonlinelibrary.com]

the friend-voice is deemed less socially or personally relevant, and the social distance between self and friend is increased.

General Discussion

Across three experiments, we asked whether an unfamiliar, externally-generated voice could be incorporated into the self and, if so, which factors modulate the resulting processing bias.

In Experiment 1, we showed that an unfamiliar voice could indeed be incorporated into the self, as measured by quicker reaction times, higher accuracy, and increased perceptual sensitivity to the new self-voice relative to a friend or stranger's voice, thus demonstrating a clear self-prioritization effect. These results are a first demonstration that a voice does not need to be self-generated in order to become integrated into our representation of self and thus be prioritized in perception as a self-relevant stimulus.

This finding is important when considering people who are unable to use their own, self-generated voice to communicate. Specifically, our results support the notion that an alternative vocal signal could quickly become personally relevant as a means of self-expression. Given that reports from AAC users often suggest that the loss of self-voice can lead to a loss of self-identity (Nathanson, 2017), ways of incorporating a new self-voice into their current representation of self are of direct importance.

In Experiment 2, we began to explore which factors might strengthen this association between the self and a new self-voice, asking whether self-prioritization would be increased when the self-voice was gender-matched to the identity of the listener.

However, we found no evidence to support our hypothesis; reaction times and perceptual sensitivity did not differ between gender-matched self-voices and non-gender-matched self-voices. Here then, gender-matching the new self-voice to the listener did not significantly moderate prioritization of that voice. Instead, a significant self-prioritization was evident in both groups for both RTs and accuracy.

Given how important voices are to representing our self (McGettigan, 2015), it might be surprising that a more personalized means of vocal expression did not generate a stronger bias, or greater prioritization. In the memory literature, for instance, Cunningham et al. (2011) suggests that the strength of an item's encompassment within the self should, intuitively, be moderated by how far the item is autobiographically consistent and by how far it matches personal tastes and goals.

Oyserman et al. (2017) too suggest that processing resources are preferentially allocated to identity-consistent stimuli since people prefer to act and make decisions according to what is identity-congruent. Within the current study, however, male and female participants showed a self-prioritization effect for a self-voice that was not – at least in gender – consistent with their self-representation. It is possible that in this context, the self-voice fulfils the same task demands regardless of gender-matching;the perceptual matching task does not require social interaction in which the new self-voice is used as a primary means of sharing or representing the self-identity. A context demanding the use of the new self-voice as a means to communicate may be required in order for a gender-matched voice to be significantly more relevant to the self than a mismatched voice. More communicative paradigms might shift the bias afforded to the self-voice from that of a merely self-owned stimulus to one of high social salience, due to the critical role it can play in the representation of self and to achieving communicative goals (McGettigan, 2015).

Two key questions remain: first, whether prioritization can ever be modulated by a stronger association between a stimulus and the self and; second, whether this association can be formed by the way in which the stimulus is encoded to the self.

In Experiment 3, we therefore asked whether the strength of the association between the self and a new voice – and so the prioritization of that voice – would be increased if the participant had made a personal choice to take ownership of it. Specifically, we examined whether listeners showed quicker reaction times and higher accuracy to the self-voice – relative to the friend-voice and stranger-voice – when the self-voice was chosen by the listener as compared to when the self-voice was randomly allocated.

Our results show that having a choice over the self-voice does interact with perceptual prioritization. However, the shift in prioritization did not manifest as quicker reaction times and increased perceptual sensitivity to the chosen self-voice relative to the unchosen self-voice. The interaction arose instead due to reduced performance for the friend-voice when the self-voice was chosen relative to when it was randomly allocated. That is, when the self-voice was chosen, the friend-voice was processed as if it were a stranger's voice, with equivalent RTs for friend and stranger. Conversely, when the self-voice was randomly allocated, there remained a distinction between friend and stranger's voices. Therefore, only when the self-voice was chosen did the distance between the self-voice and the other two voices increase, such that the self-voice became more prominent because of reduced prioritization of the other voices.

The precise manifestation of the effect of choice in Experiment 3 is intriguing. While clearly indicating that choice increased the perceptual separation of the self-voice from the other identities, this seems to have happened via an 'othering' of the friend voice rather than enhanced prioritization of the self-voice per se. As there is no prior report of the effect of choice on self-prioritization within a perceptual matching paradigm, we cannot tell if the effect observed here might be particular to voices or indeed would be found using any category of stimuli (e.g., visual objects). It will be important to address the generalizability of the finding in future work. It will also be relevant to explore the interplay of mere ownership versus agency (i.e., choosing to take ownership) in this paradigm. Using different tasks, previous studies have shown that self-bias is most reliably observed when items are both owned and chosen by the participant, but remains even if those items are chosen for another person (Huang et al., 2009). The current paradigm with voices could be extended with a manipulation of *which* voice is chosen, for example with some participants selecting the voice to represent themselves, and some choosing the friend-voice.

In our study, we have begun to qualify which factors could be harnessed to aid the incorporation of a new, externally-generated voice within the self. We have shown that a new voice can be incorporated into the self after only brief exposure, as evidenced by a self-prioritization effect in all three experiments. Moreover, personal choice in picking the new self-voice may modulate the distinctiveness of that voice relative to others. Finally, although matching the new self-voice to the gender identity of the listener did not seem to modulate prioritization of it here, it is possible that the absence of an effect is due to the lack of a need to use this voice to represent oneself socially within the task. Given that the voice is a highly salient social stimulus, a more immersive, social interaction-based paradigm, in which the new self-voice is actively used to achieve communicative goals, could elicit a stronger bias than we have shown here. In such a context, prioritization may be modulated by how closely the vocal stimulus can represent the self, including the gender identity of the speaker.

The current results support the case for more personalized voices in which the speaker has agency in choosing their new auditory identity. Future studies should utilize more immersive communicative paradigms to more closely model the adoption of new voices in patients. These studies could better measure the integration of a new voice into the self, the factors that may facilitate integration, and the subsequent bias that is afforded to a new self-voice as a primary means of representing self-identity. These proposed future studies can proceed, importantly, resting on the evidence presented here for the first time; it is possible to incorporate a new voice into our self-identity that would otherwise have been perceived as belonging to an 'other'. We can do this after only brief exposure to it and can enhance our bias towards it through making a personal choice to own it.

Acknowledgements

Conflict of interest

All authors declare no conflict of interest

Author contributions

Bryony Payne (Conceptualization; Formal analysis; Investigation; Resources; Writing –original draft; Writing – review & editing); Nadine Lavan (Formal analysis; Writing – original draft; Writing – review & editing); Sarah Knight (Resources); Carolyn McGettigan (Conceptualization; Funding acquisition; Writing – original draft; Writing – review & editing).

Data availability statement

Data and materials for this study can be accessed via the web links below: Experiment 1: https://osf.io/hg3w6/; Experiment 2: https://osf.io/r6g2m/; Experiment 3: https://osf.io/wz2am/.

References

Allan, K., Morson, S., Dixon, S., Martin, D., & Cunningham, S. (2017). Simulation-based mentalizing generates a "proxy" self-reference effect in memory. *Quarterly Journal of Experimental Psychology*, *70*, 1074–1084. https://doi.org/10.1080/17470218.2016.1209532

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2019). Gorilla in our Midst: Gorilla.sc, a new web-based experiment builder. *Behavior Research Methods*, *52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Audacity Team (2019). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.3.2 Retrieved March 2019 from https://audacityteam.org/.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *Journal of Statistical Software*, *67*, 1–23. https://doi.org/10.18637/jss.v067.i01

Benoit, R.G., Gilbert, S.J., Volle, E., & Burgess, P.W. (2010). When I think about me and simulate you: medial rostral prefrontal cortex and self-referential processes. *NeuroImage*, *50*, 1340–1349. https://doi.org/10.1016/j.neuroimage.2009.12.091

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.26, retrieved 5 October 2019 from http://www.praat.org

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Cloutier, J., & Macrae, C. (2008). The feeling of choosing: Self-involvement and the cognitive status of things past. *Consciousness and Cognition*, *17*(1), 125–135. https://doi.org/10.1016/j.concog.2007.05.010

Cunningham, S., Brady-Van den Bos, M., & Turk, D. (2011). Exploring the effects of ownership and choice on self-memory biases. *Memory*, *19*, 449–461. https://doi.org/10.1080/09658211.2011.584388

Cunningham, S.J., Turk, D.J., Macdonald, L.M., & Macrae, C.N. (2008). Yours or mine? Ownership and memory. *Consciousness and Cognition*, *17*, 312–318. https://doi.org/10.1016/j.concog.2007.04.003

Gregg, A., Mahadevan, N., & Sedikides, C. (2017). The SPOT effect: people spontaneously prefer their own theories. *Quarterly Journal of Experimental Psychology*, *70*, 996–1010. https://doi.org/10.1080/17470218.2015.1099162

Holzman, P.S., & Rousey, C. (1966). The voice as a percept. *Journal of Personality and Social Psychology*, *4*(1), 79–86. https://doi.org/10.1037/h0023518

Huang, Y., Wang, L., & Shi, J. (2009). When do objects become more attractive? The individual and interactive effects of choice and ownership on object evaluation. *Personality and Social Psychology Bulletin*, *35*, 713–722. https://doi.org/10.1177/0146167209333046

Hughes, S., & Harrison, M. (2013). I like my voice better: self-enhancement bias in perceptions of voice attractiveness. *Perception*, *42*, 941–949. https://doi.org/10.1068/p7526

Hughes, S.M., Mogilski, J.K., & Harrison, M.A. (2014). The perception and parameters of intentional voice manipulation. *Journal of Nonverbal Behavior*, *38*(1), 107–127. https://doi.org/10.1007/s10919-013-0163-z

Jones, J., Pelham, B., Carvallo, M., & Mirenberg, M. (2004). How do I love thee? Let me count the Js: Implicit egotism and interpersonal attraction. *Journal of Personality and Social Psychology*, *87*, 665–683. https://doi.org/10.1037/0022-3514.87.5.665

Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*, Chichester: Wiley-Blackwell. https://doi.org/10.1002/9781444395068

Lass, N.J., Hughes, K.R., Bowyer, M.D., Waters, L.T., & Bourne, V.T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *The Journal of the Acoustical Society of America*, *59*, 675–678.

Lenth, R. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

McGettigan, C. (2015). The social life of voices: studying the neural bases for the expression and perception of the self and others during spoken communication. *Frontiers in Human Neuroscience*, *9*(129), 1–4. https://doi.org/10.3389/fnhum.2015.00129

McGettigan, C., Eisner, F., Agnew, Z.K., Manly, T., Wisbey, D., & Scott, S.K. (2013). T'ain't what you say, it's the way that you say it—left insula and inferior frontal cortex work in interaction with superior temporal regions to control the performance of vocal impersonations. *Journal of Cognitive Neuroscience*, *25*, 1875–1886. https://doi.org/10.1162/jocn_a_00427

McGettigan, C., & Scott, S.K. (2014). Voluntary and involuntary processes affect the production of verbal and non-verbal signals by the human voice. *Behavioral and Brain Sciences*, *37*, 564–565. https://doi.org/10.1017/S0140525X13004123

Nathanson, E. (2017). Native voice, self-concept and the moral case for personalized voice technology. *Disability and Rehabilitation*, *39*(1), 73–81. https://doi.org/10.3109/09638288.2016.1139193

Owren, M., Berkowitz, M., & Bachorowski, J. (2007). Listeners judge talker sex more efficiently from male than from female vowels. *Perception & Psychophysics*, *69*, 930–941. https://doi.org/10.3758/bf03193930

Oyserman, D., Lewis, N., Yan, V., Fisher, O., O'Donnell, S., & Horowitz, E. (2017). An identity-based motivation framework for self-regulation. *Psychological Inquiry*, *28*, 139–147. https://doi.org/10.1080/1047840x.2017.1337406

Payne, S., Tsakiris, M., & Maister, L. (2017). Can the self become another? Investigating the effects of self-association with a new facial identity. *Quarterly Journal of Experimental Psychology*, *70*, 1085–1097. https://doi.org/10.1080/17470218.2015.1137329

Pohlert, T. (2014). The pairwise comparison of mean ranks package (PMCMR). R Package. Retrieved from https://CRAN.R-project.org/package=PMCMR .

Puts, D., Hill, A., Bailey, D., Walker, R., Rendall, D., Wheatley, J. . . . Ramos-Fernandez, G. (2016). Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1829), 20152830. https://doi.org/10.1098/rspb.2015.2830

R Core Team (2013). *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/ .

Schäfer, S., Wesslein, A., Spence, C., Wentura, D., & Frings, C. (2016). Self-prioritization in vision, audition, and touch. *Experimental Brain Research*, *234*, 2141–2150. https://doi.org/10.1007/s00221-016-4616-6

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers*, *31*, 137–149. https://doi.org/10.3758/BF03207704

Sui, J., He, X., & Humphreys, G.W. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1105–1117. https://doi.org/10.1037/a0029792

Sui, J., & Humphreys, G. (2017). Aging enhances cognitive biases to friends but not the self. *Psychonomic Bulletin & Review*, *24*, 2021–2030. https://doi.org/10.3758/s13423-017-1264-1

Sui, J., Rotshtein, P., & Humphreys, G.W. (2013). Coupling social attention to the self forms a network for personal significance. *Proceedings of the National Academy of Sciences USA*, *110*, 7607–7612. https://doi.org/10.1073/pnas.1221862110

Symons, C.S., & Johnson, B.T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*, 371–394. https://doi.org/10.1037/0033-2909.121.3.371

Woods, K.J., Siegel, M.H., Traer, J., & McDermott, J.H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, and Psychophysics*, *79*, 2064–2072. https://doi.org/10.3758/s13414-017-1361-2