Empirical Research Paper

Multimodal Cues to Change Your Mind: The Intertwining of Faces, Voices, and Behaviors in Impression Updating

Personality and Social Psychology Bulletin
1–14

DOI: 10.1177/01461672241273209
journals.sagepub.com/home/pspb
Sage

Matteo Masi[1,2], Simone Mattavelli[1,3], Fabio Fasoli[2], and Marco Brambilla[1]

Abstract

Impressions of others are formed from multiple cues, including facial features, vocal tone, and behavioral descriptions, and may be subject to multimodal updating. Four experiments ($N = 803$) examined the influence of a target's face or voice on impression updating. Experiments 1a-1b examined whether behavior-based impressions are susceptible to updating by incongruent information conveyed by the target's face, voice, or behavior (within-participant manipulation). Both faces and voices updated impressions with comparable strength, but less than behaviors. Experiment 2, contrasting faces and voices only (between-participants manipulation), showed that voices outperformed faces regardless of how impressions were formed (i.e., via behavioral vs. nonbehavioral information). Experiment 3 found no difference when comparing faces and voices in a within-participant design and controlling for stimulus attractiveness. Our work highlights the importance of multimodal cues for impression updating and shows that the relative power of faces and voices depends on contextual factors.

Keywords

impression updating, face, voice, behavior, multimodal perception

Received April 18, 2023; revision accepted June 27, 2024

Imagine that John has just moved into your neighborhood, and you have not met him yet. A friend of yours told you that John used to be a bully in high school. From this information, you might think of John as someone who should not be trusted. A few days later, you spot him in his front yard. Think of two alternative situations. In one case, you see only his face from a distance. In the other case, you only hear his voice while talking on the phone. In both cases, John's face and voice are more pleasant than expected and create impressions that do not match that based on John's past behavior (i.e., bullying others). This example suggests that social interactions can be shaped by information derived from others' behaviors, as well as faces and voices, and that such information integrates over time. Here, we tested whether facial and vocal cues update first impressions formed on verbal descriptions of target persons' past behaviors.

Impressions From Behaviors, Faces, and Voices

Understanding others is essential in coping with the social world (Fiske et al., 1992), and various cues can be sources of social impressions (Freeman et al., 2020; McArthur & Baron, 1983). People often base their impressions on others' behaviors (Skowronski et al., 2008; Uleman & Kressel, 2013). Getting to know that someone engaged in a specific behavior is sufficient to evaluate the person (Schneid et al., 2015) and infer their states (Kruse & Degner, 2021) and traits (Cone et al., 2017). Indeed, behaviors are often used to make causal inferences about a person's disposition (e.g., Paul kicked a dog, and he did it because he is aggressive, Newman, 1996) that are automatically stored in the memory (Todorov & Uleman, 2002). This occurs because people's past behaviors shape our expectations concerning future actions, which affects perceivers' interactions with them (Snyder & Stukas, 1999; Vonk, 1994). For instance, knowing that Paul once kicked a dog would make people reluctant to let him dog-sitting (e.g., McCarthy & Skowronski, 2011).

Just like observing/being told about someone's deeds is a way to form an impression, listening to a person's voice or

[1]University of Milano-Bicocca, Milan, Italy
[2]University of Surrey, Guildford, UK
[3]Vita-Salute San Raffaele University, Milan, Italy

Corresponding Author:
Matteo Masi, Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, Milano 20126, Italy.
Email: matteo.masi@unimib.it

seeing their face conveys information that is difficult to ignore. Nonbehavioral cues are regularly evaluated to form impressions (Ambady et al., 2000; McAleer & Belin, 2018; Todorov et al., 2015). Specifically, a glance at a face conveys an impressive amount of information (Sutherland & Young, 2022; Zebrowitz, 2017). Notwithstanding people's ability to infer social category membership (e.g., ethnicity, age, gender, sexual orientation; Johnson et al., 2015; Rule et al., 2015; Rule & Sutherland, 2017), the face is also perceived as a cue of others' dispositions (Foo et al., 2021; Todorov et al., 2015). In a few milliseconds, people can form a face-based impression of others which, in turn, influences social interactions (e.g., Olivola & Todorov, 2010; Willis & Todorov, 2006). Similarly, listening to a few syllables from a speaker's voice makes a listener infer several speakers' characteristics (McAleer & Belin, 2018). Indeed, the voice is used to infer speakers' social group membership (e.g., ethnicity, age, gender, sexual orientation; Gluszek & Dovidio, 2010; Latinus & Belin, 2011; Masi & Fasoli, 2022). Moreover, listeners can infer speakers' dispositions after a few milliseconds of voice exposure (Mileva & Lavan, 2023) that influence social behaviors (e.g., Klofstad et al., 2012).

In summary, individuals rely on others' behaviors, faces, or voices as informative cues to shape their impressions. Social perception theories (e.g., Freeman et al., 2020; McArthur & Baron, 1983) highlight the importance of integrating information from multiple modalities to effectively navigate social interactions. Virtually, any instance of social information might contribute to our understanding of others' dispositions, ultimately aiding in predicting their intentions. Our perceptions of others can be continuously updated, especially when we are confronted with new information that contradicts previous impressions (Cone et al., 2017). Despite its relevance, the interplay between multimodal cues in the dynamic process of impression formation and updating has received scant attention thus far.

Impression Updating With Behavioral and Nonbehavioral Information

People are motivated to solve inconsistencies between multiple information (Asch & Zukier, 1984). Thus, under specific conditions, new information is used to update the original impression of the target person (Cone et al., 2017; Moskowitz et al., 2022). For instance, Cone and Ferguson (2015) showed that impressions based on a series of positive behaviors performed by a target are reversed when knowing that the same person performed a single (and diagnostic) countervailing behavior (e.g., molesting a child). Most work on impression updating employed verbal descriptions of behaviors as cues for both impression formation and impression updating (see Cone et al., 2017 for a review). Only recently, studies have started looking at how behavioral and nonbehavioral information interact in the updating process. Shen et al. (2020) demonstrated that an impression formed upon a face can be updated by additional and countervailing behavioral information (see also Shen et al., 2023). Yet, Shen and colleagues considered only facial information as a nonbehavioral cue of impressions. In addition, they examined the power of behavioral cues to update such impressions, ignoring the other way around. Hence, the single type of nonbehavioral cues considered in their study (i.e., faces) and the unidirectionality of the updating effect (i.e., from nonbehavioral to behavioral cues) limits the generalization of these findings to alternative instances of multimodal impression updating, calling for a throughout investigation of the other possible combinations of cues. For instance, whether the addition of facial or vocal cues significantly change the initial impression formed when observing a behavior remains an unanswered question.

That nonbehavioral information can alter impressions formed from someone's actions would highlight how much people value judgments based on these cues. One might argue that nonbehavioral cues carry more ambiguous information about a person than their actual behaviors. For example, hearing a pleasant voice might leave more room for uncertainty about a person's qualities than hearing about someone's misdeeds. Yet, people might still be inclined to rethink their behavior-based impressions when they hear others' voices or see others' faces. In line with this idea, research on impression formation showed that judgments based on (diagnostic) behavioral information are influenced by the co-occurrence of incongruent facial (Rezlescu et al., 2012) or vocal information (Ko et al., 2009). However, this possibility is still to be tested in impression updating, that is, when behavioral and nonbehavioral information is processed in sequence rather than in parallel.

When examining multimodal impressions, an important inquiry pertains to the relative power of different cues in shaping our evaluations. This question has been explored from various angles, considering criteria such as identity and affect recognition (De Gelder & Bertelson, 2003; Young et al., 2020). Several studies delved into the differences between faces and voices in their importance for impression formation (e.g., Hansen et al., 2017; Mileva et al., 2018; Rezlescu et al., 2015; Zuckerman & Sinicropi, 2011). These investigations suggested that such cues are integrated to form overall impressions, with effects that can be either additive or multiplicative. Furthermore, the relative importance of faces and voices varies depending on the trait being observed. For example, in simultaneous impressions, perceptions of trustworthiness tend to be more strongly influenced by voices than by faces (but only when certain criteria are respected, Vives et al., 2023; cf. Mileva et al., 2018; Rezlescu et al., 2015), whereas for other social judgments (e.g., attractiveness) facial cues may be more relevant (Liu et al., 2023).

Studies on impression formation have also shown that, despite some similarities, face- and voice-based personality

perceptions differ, with the voice eliciting more positive impressions than the face on judgments of capability, approachability, and reliability (Jiang et al., 2024). Other studies focusing on the consequences of social categorization and impression formation showed that the voice exerts a stronger influence than the face in some decision-making processes, such as hiring decisions (Fasoli et al., 2017; Rakić et al., 2011) or voting behavior (Mileva et al., 2020). These findings seem to show that on certain occasions the voice is a more diagnostic cue for social judgments and their consequences. However, such findings are confined to scenarios where cues are presented and compared in isolation or simultaneously, failing to illuminate the dynamics of sequential cue presentation that characterizes impression updating. In everyday interactions, information may be provided sequentially rather than simultaneously, such as when we see or hear a job applicant after reading their resumé, or when we first read about a political candidate in the newspaper and then hear their voice on the radio or see their face on social media.

To gain insights into the differences between faces and voices in updating power, it is essential to examine existing research findings on impression updating based solely on nonbehavioral cues. Masi et al. (2022) compared the updating capabilities of faces and voices in cross-modal impressions and investigated whether positive and negative information from one cue (e.g., the face) could modify impressions formed from another cue (e.g., the voice). By examining this updating effect bidirectionally, they found that voices had a greater influence in updating face-based impressions, particularly negative ones, compared with the reverse scenario. These findings parallel Hansen et al.'s (2018), albeit in the context of ethnicity-based cues, where voices (i.e., accent) were found to have more influence in updating cross-modal impressions than faces (i.e., ethnicity facial cues). Although limited, this emerging evidence suggests that individuals may assign greater significance to the voice when it comes as a second cue in a sequential presentation, while the face is assigned less weight.

There might be multiple reasons why voices could outperform faces in updating impressions. For example, research on emotion perception suggests that the voice is a type of information that is perceived as richer than the face, consisting of several paralinguistic (e.g., pitch, speed, volume) and linguistic features that may attract a listener's attention to a greater extent, resulting also in increased emphatic accuracy (Kraus, 2017). Thus, what might distinguish the two nonbehavioral cues for impression updating is that a voice forces attention to, and a deeper elaboration of, social information. This may also be related to the belief that when a person speaks to someone, it is often because the person intends to do so, whereas a person's face is often visible without any intention to show it (Lavan et al., 2021). Hence, it may be that people value information from the voice as more informative because it is perceived as more "agentic" compared with the face. Thus, voices might represent a richer source of information capable of generating inferences about a person's intentionality. This might be extremely relevant for impression updating, which requires the perceiver to "change their mind" about the target. In other words, it might be easier to revise an initial impression of a target when a countervailing piece of information comes from a source that allows inferences on the target's intentionality. Ultimately, the voice might be more diagnostic of a person's true nature (as also suggested by Masi et al., 2022). Here, we explore whether a superior influence of voices over faces is confirmed in the context of impression updating.

The Present Research

The present research examined multimodal impression updating and aimed to shed light on the influence of faces and voices in this process. Previous studies on impression updating have predominantly focused on verbal descriptions of behavior and have, thus, overlooked the role of nonbehavioral cues in shaping impressions over time. We aimed to fill this gap by examining how facial and vocal cues contribute to the updating of impressions formed from the description of a person's behavior. Based on previous research on nonbehavioral impression updating (which did not include behavioral information), we expected voices to exert a stronger updating effect than faces, possibly because the voice enhances attention to and elaboration of the social information more than the face (e.g., Hansen et al., 2018; Kraus, 2017; Masi et al., 2022).

In our studies, we decided to test only the negative to positive impression updating route. There is a well-established valence asymmetry in people's willingness to revise the target's negative behavioral impression in light of newly acquired positive information, especially when it pertains to morality/trustworthiness, as opposed to the other way around (Cone et al., 2017; see also Baumeister et al., 2001). As our main goal was to study whether multimodal cues can update first impressions, especially those based on behavioral cues, focusing on the most difficult route to updating (i.e., from negative to positive) would yield the most insightful results (see Shen et al., 2020 for a similar reasoning).

Experiments 1a-1b tested whether positive (i.e., trustworthy) faces and voices could update first impressions of targets based on their negative (i.e., untrustworthy) behaviors and whether any difference emerged between the two cues. We compared the relative impact of the target's face and voice in impression updating with the impact of positive (i.e., trustworthy) behavior to update impressions. In Experiment 2, we assessed the relative impact of faces and voices in updating impressions based on behavioral or nonbehavioral information (i.e., face-to-voice or voice-to-face updating). Experiment

3 was a preregistered investigation (https://osf.io/w38bx/) conducted to replicate Experiments 1a-1b through controlling for additional factors (i.e., design shortcomings, cues attractiveness).

All studies received approval from the local university's ethics committee. All measures, manipulations, and exclusions in the studies are disclosed. Materials, data, and analysis codes were uploaded to Open Science Framework (https://osf.io/de3rp/). The preregistered analyses are reported in the manuscript or Supplemental Materials, and any deviations from the original plan are indicated.

Experiments 1a-1b

In Experiment 1a, we tested whether a negative impression of a male target formed upon a negative behavior could be updated by positive information conveyed by his face or voice. We also examined whether presenting additional positive behavioral information led to impression updating as in previous studies (Cone et al., 2017). This was also done to assess the relative power of faces and voices compared with a unimodal cue of information. Based on the superior role played by voices over faces in previous research on impression updating (Hansen et al., 2018; Masi et al., 2022), we hypothesized that voices would lead to stronger impression updating than faces. As unimodal (behavior-to-behavior) and multimodal (behavior-to-face, behavior-to-voice) impression updating had not yet been compared, we explored their differences. We ran Experiment 1b to replicate Experiment 1a's findings as our experiment was the first using this design.

Method

Design. We employed a 3 (Second information type: Face vs. Voice vs. Behavior) × 2 (Measurement time: Time1 vs. Time2) within-subjects design. The first information was always negative, and the second information was always positive. Impressions measured at Time 1 allowed us to assess impression formation effects. Critical for the present investigation was the interaction of the two factors accounting for the impact of second information types on impression updating and their comparison.

Participants and Power Analysis. Due to the novelty of the experimental design, we could not rely on any prior estimated effect size. Therefore, considering our economic resources (Lakens, 2022), we aimed to recruit at least 100 participants in the first experiment and approximately the same number of participants in the replication. Thus, in Experiment 1a the sample consisted of 122 Italian participants (58 females, $M_{age} = 24.15$, $SD_{age} = 2.61$) while Experiment 1b involved 113 Italian participants (48 females, $M_{age} = 26.92$, $SD_{age} = 7.07$).

To assess statistical power in the absence of sample size determination for linear mixed models, we conducted a safeguard sensitivity analysis (Perugini et al., 2014). This analysis tells us if the experiments had adequate power (> 80%) to detect the crucial comparisons when considering the uncertainty of the estimated effects (i.e., CI). Using the "simr" R package (Green & MacLeod, 2016). We simulated the power of each experiment after 1,000 repetitions when substituting observed effects with the lower 20th percentile (Perugini et al., 2014). We did so for the critical comparisons, that is, ability to do updating of each cue as well as the differences in their updating power. We did not conduct safeguard analysis for nonsignificant differences. Experiment 1a was sufficiently powered to detect the lower bounds of the updating effect of each type of cue (power > 99%) as well as the behavior vs. face and behavior vs. voice effects (power > 98%). Experiment 1b was powered enough to detect the lower bounds of the updating effect of each type of cue (power > 99%) but had low power for the lower bounds of the behavior versus face (power = 28%) and behavior versus voice effects (power = 52%). Detailed results are available as Supplementary Materials.

Materials. As in previous research (Shen et al., 2020), cues were manipulated in trustworthiness. Trustworthiness is the leading dispositional trait for impression formation, and judgments on this dimension approach valence-based ones (Brambilla et al., 2021; McAleer et al., 2014; Todorov et al., 2015). For all the cues (i.e., behaviors, faces, and voices), we identified and selected those with a comparable moderate level of perceived trustworthiness (±1 on a −3 = untrustworthy to +3 = trustworthy scale). By balancing cues on trustworthiness, we attempted to ensure that any differences observed in the resulting impressions were driven solely by the type of cue (see Brambilla et al., 2019).

The materials underwent thorough pretesting, the details of which are provided in the Supplementary Materials. We used the pretest trustworthiness ratings and selected six untrustworthy behaviors ($M = -0.99$, $SD = 0.16$) for initial impressions, and two trustworthy behaviors ($M = 1.16$, $SD = 0.05$) for subsequent countervailing information. In addition, as trustworthy nonbehavioral cues, we chose the voices of two male speakers uttering a 3000 ms neutral-in-valence sentence ($M = 1.05$, $SD = 0.19$), and two digitally generated male faces, created using AI technology, depicted in a frontal pose ($M = 0.85$, $SD = 0.26$).

Procedure. Participants were recruited via Prolific and rewarded with £9.00/hr for their participation (Experiment 1a: £0.63, 4 min; Experiment 1b: £0.78, 5 min). Participants were introduced to an online experiment on impression formation implemented on Qualtrics. They were asked to self-isolate, check their headphones' functioning, and silence other devices. They were told that their goal was to evaluate three targets whose characterizing information were going to be provided.[1] Hence, three targets were presented in random sequence, one after the other. Throughout the experiment,

this sequence was kept constant in each part. Each target's name was shown in the upper part of the screen, along with a statement presenting a negative behavior the target allegedly performed and randomly extracted from the set of six negative behaviors. The statement stayed on the screen for about 3000 ms to match the length of the voice recordings. Afterward, the first impression of each target was measured: Participants answered the question, "Your global impressions of [name] is?" on a scale ranging from −3 (*negative*) to + 3 (*positive*) (Time 1) in the same order of presentation as before. Next, for each target, participants received countervailing information. Namely, each target's name was presented again in the upper part of the screen, alone and in the same order as previously presented, and was randomly paired with a voice, a face, or another statement describing a past behavior performed by the target. Next, targets were reevaluated on the same global impression scale employed in the first section (Time 2). Participants were then debriefed and thanked.

Data Preparation and Analysis. In all experiments, data were analyzed in R Studio. We analyzed the results in a linear mixed model (package *lmerTest*, Kuznetsova et al., 2017). In Experiments 1a-1b, second information type, measurement time, and their interaction were included as fixed effects. We attempted to maximize the random effect's structure avoiding issues of convergence/singularity following Barr et al. (2013). We included a by-subject random slope accounting for measurement time and a by-item random intercept for the combination of the pairs of cues. More complex models with random slopes failed to converge or resulted in singularity warnings. Significant interactions were inspected with simple effects analysis (p values were "Holm"-adjusted when required, package *emmeans*, Lenth, 2022).[2]

Results

Experiment 1a. Descriptive statistics are reported in Table 1 and results in Figure 1. We found a main effect of time, $b = -0.77$, $SE = .05$, $t(121) = -16.04$, $p < .001$, meaning that impressions were significantly updated irrespective of the second information type. We also found an effect of information type, $F(2, 27.24) = 7.70$, $p = .002$, that was qualified by a significant interaction, $F(2, 452.83) = 15.43$, $p < .001$. Direct contrasts inspecting the interaction revealed that behavioral first impressions of the three targets were comparably negative at Time 1, $b < 0.04$, $SE = .16$, $t(60) < 0.25$, $p > .967$. Moreover, all types of second information were able to update first impressions significantly, behavior: $b = -2.17$, $SE = .15$, $t(448.57) = -14.70$, $p < .001$; face: $b = -1.24$, $SE = .15$, $t(448.57) = -8.43$, $p < .001$; voice: $b = -1.24$, $SE = .15$, $t(448.57) = -8.38$, $p < .001$. Comparing the differences in updating due to each type of second information (i.e., Time 2 – Time 1) indicated that behavioral information was more effective in updating first impressions than faces, $b = 0.93$, $SE = .19$, $t(459.78) = 4.79$, $p < .001$, and voices, $b = 0.93$, $SE = .19$, $t(459.78) = 4.83$, $p < .001$, while faces and voices did not differ from each other, $b = 0.008$, $SE = .19$, $t(459.78) = -0.04$, $p = .966$.

Table 1. Experiment 1a.

	Time 1		Time 2		Score	
Second information	*M*	*SD*	*M*	*SD*	*M*	*SD*
Behavior	−0.84	1.19	1.34	1.15	2.17	1.49
Face	−0.83	1.35	0.42	1.13	1.25	1.70
Voice	−0.79	1.26	0.44	1.23	1.24	1.59

Note. Means and standard deviations of the evaluation given to the targets as a function of time (Time 1 vs. Time 2) and the second information type (Behavior vs. Face vs. Voice). The column Score represents the difference between the evaluations given at the two stages (Time 2 – Time 1). Judgments are reported on the original scale ranging from −3 (negative) to 3 (positive).

Experiment 1b. Descriptive statistics are reported in Table 2 and results in Figure 1. We found a main effect of time, $b = -0.72$, $SE = .04$, $t(112) = -16.30$, $p < .001$, meaning that impressions were significantly updated irrespective of the second information type. We also found an effect of information type, $F(2, 24.20) = 8.87$, $p = .001$, that was qualified by a significant interaction effect, $F(2, 418.35) = 4.43$, $p = .012$. Direct contrasts inspecting the interaction first revealed that behavioral first impressions of the three targets were comparably negative at Time 1, $b < 0.27$, $SE = .16$, $t(50) < 1.67$, $p > .224$. Moreover, all types of second information were able to update first impressions significantly, behavior: $b = -1.74$, $SE = .13$, $t(414.45) = -12.86$, $p < .001$; face: $b = -1.35$, $SE = .13$, $t(414.45) = -9.99$, $p < .001$; voice: $b = -1.24$, $SE = .13$, $t(414.45) = -9.14$, $p < .001$. By comparing the differences in updating due to each type of second information (i.e., Time 2 – Time 1), we found that behavioral information was more effective in updating first impressions than voices, $b = 0.50$, $SE = .18$, $t(420.58) = 2.84$, $p = .014$, while the difference with faces was not significant, $b = 0.40$, $SE = .18$, $t(420.58) = 2.19$, $p = .058$. The difference between faces and voices in updating score was not significant, $b = 0.11$, $SE = .18$, $t(420.58) = 0.65$, $p = .517$.

Discussion

Results from Experiments 1a-1b were informative in three different ways. First, we found that behavior-based negative impressions can be updated in both a unimodal and multimodal fashion. Indeed, target cues for which a countervailing face or voice was presented at Time 2 were judged more positive than after initial exposure to the negative behavioral

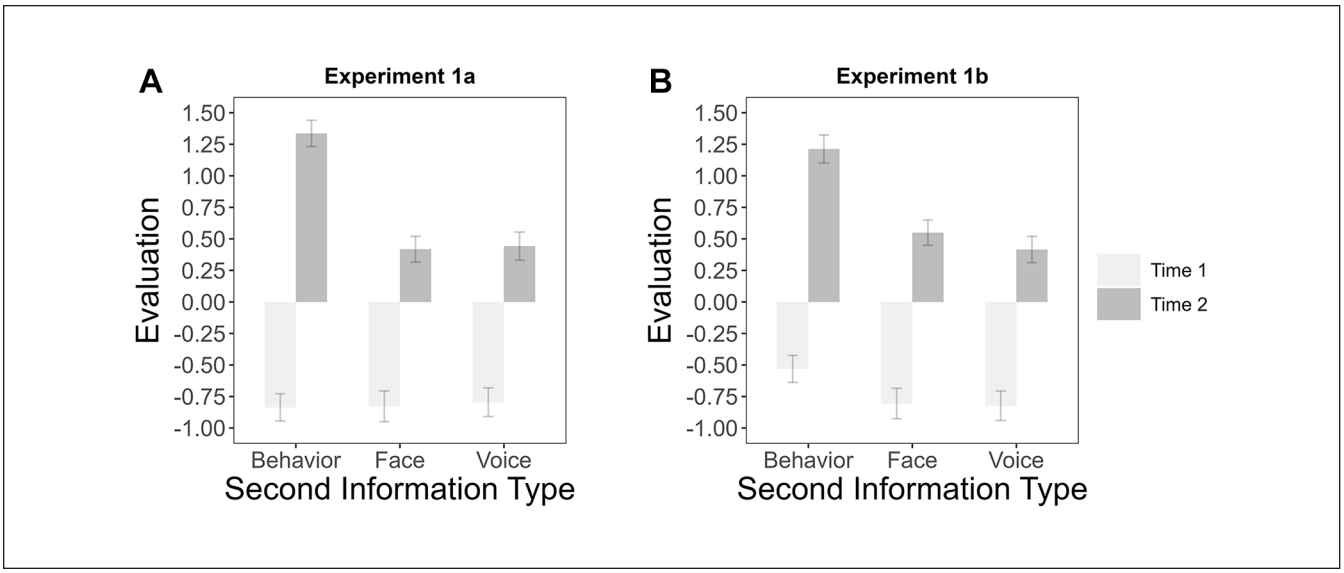

Figure 1. Bar Graph of Experiments 1a-1b Results With 95% Confidence Intervals.
Note. Second information type (Behavior vs. Non-behavior) is depicted on the x-axis. The light gray bar represents the Time 1 evaluation, and the dark gray bar represents the Time 2 evaluation.

Table 2. Experiment 1b.

	Time 1		Time 2		Score	
Second information	*M*	*SD*	*M*	*SD*	*M*	*SD*
Behavior	−0.53	1.14	1.21	1.18	1.74	1.49
Face	−0.80	1.29	0.55	1.07	1.35	1.30
Voice	−0.82	1.24	0.42	1.11	1.24	1.45

Note. Means and standard deviations of the evaluation given to the targets as a function of time (Time 1 vs. Time 2) and second information type (Behavior vs. Face vs. Voice). The column Score represents the difference between the evaluations given at the two stages (Time 2 – Time 1). Judgments are reported on the original scale ranging from −3 (negative) to 3 (positive).

information. Second, behavioral information was the most effective in updating behavior-based impressions. Third, and not in line with our hypothesis, no evidence for a difference in impression updating between faces and voices was found. This latter finding diverges from previous evidence showing a superior updating influence of voices over faces (Masi et al., 2022).

The nonsignificant difference between faces and voices in impression updating based on behavioral information might reflect a genuine comparable power of the two nonverbal cues. Alternatively, it could result from a specific feature of the design used in the two experiments. Because faces and voices were directly compared with a more diagnostic cue (i.e., behaviors), participants might have mentally aggregated the comparison into behavioral vs. nonbehavioral information. Such a reframing might have contributed to reducing the discrepancy between the vocal and the facial information. Experiment 2 was designed to tackle this issue.

Experiment 2

Experiments 1a-1b showed that faces and voices were equally effective in updating first impressions based on the target's untrustworthy behavior. However, previous studies have shown that faces and voices update cross-modal impressions differently (Masi et al., 2022). In those studies, positive voices updated negative impressions generated from the target's face to a greater extent than the other way around. Hence, based on Experiments 1a-1b, one could argue that faces and voices differ in their updating effectiveness only when impressions are generated from the alternative cues (i.e., voices are more effective in updating face-based impressions than faces are in updating voice-based impressions) but that they are comparable when it comes to updating behavior-based impressions. Experiment 2 tested this possibility empirically by comparing (positive) faces and voices in updating (negative) impressions generated either from the behavioral information or the nonbehavioral alternative channel. Importantly, we also wanted to ensure that the comparison between the influence of faces and voices was not affected by the comparison with a likely more diagnostic cue. Hence, we removed the behavioral information at Time 2. Moreover, while in Experiments 1a-1b the type of second information was manipulated within participant, here we manipulated this variable between participants.

As this experiment also required untrustworthy faces (and voices) as cues for an impression at the first stage and finding them among those created via AI and employed in Experiments 1a-1b was no trivial task, we used real faces extracted from a high-quality photographs database (Chicago Face Database; Ma et al., 2015). This increased the

ecological validity of our work and allowed us to replicate cross-modal impression updating findings with real photographs instead of computer-generated avatar faces (see Masi et al., 2022).

Method

Design. We employed a 2 (First information type: Behavior vs. Non-behavior) × 2 (Second information type: Face vs. Voice) × 2 (Measurement Time: Time 1 vs. Time 2) mixed design, with the second factor manipulated between-subjects. In this way, we could directly observe whether voices or faces differ in their updating impact on impressions based on behavioral and nonbehavioral information. The first information was always negative while the second information was always positive.

Participants and Power Analysis. As no prior experiment used a similar design, we had no reference for effect sizes to consider in an a priori power analysis. We aimed to collect at least 350 participants. The final sample comprised 358 participants (M_{age} = 28.44, SD_{age} = 8.55, 191 females).

We conducted a safeguard analysis to show that the experiments had sufficient power to detect the critical effect. The results showed that we had sufficient power for detecting the lower bound of the updating effect of both faces and voices on first impressions (power > 99%) and the face vs. voice difference (power = 89%).

Materials. We tested a new set of cues because this experiment required us to employ untrustworthy faces and voices as cues for first impressions (see Supplementary Materials). Five behaviors were selected to be untrustworthy (M = −1.33, SD = 1.31). We also selected five male speakers evaluated as trustworthy (M = 1.26, SD = 1.21) and five as untrustworthy (M = −1.08, SD = 1.36). Then, we extracted pictures of male targets from the Chicago Face database, five of which were perceived as trustworthy (M = 1.26, SD = 1.26) and five as untrustworthy (M = 0.95, SD = 1.28).

Procedure. Participants were recruited via Prolific Academic (rewarded £8/hr, £0.59 for a 4-min experiment) and introduced to an online impression formation experiment implemented on Qualtrics. First, they were exposed to two target individuals, presented one at a time and in a random order, with their names written in the upper part of the screen. One target name was associated with the description of a negative behavior randomly extracted from the pool and shown at the center of the screen, and the other target with a negative nonbehavioral information randomly extracted as well. For half of the participants, the first nonbehavioral information was the target's face, while for the other half the first information was the target's voice. Then, participants were asked to indicate their impression of each target presented in the same order as before on a 7-point scale (−3 = negative, +3 = positive) as in Experiments 1a-1b (Time 1). In the second part, depending on the condition, participants were presented either with the targets' face or voice (both positive), randomly paired with the two targets. To illustrate, at Time 1, Participant x was initially presented with an individual (e.g., Paolo) associated with a negative past behavior and another individual (e.g., Diego) associated with a negative face (vs. voice, for participant y). Subsequently, at Time 2, Participant x was exposed to Paolo's and Diego's positive voices (vs. face, for participant y). Then, a second evaluation was asked on the same scale (Time 2). Participants were finally debriefed and thanked.

Data Preparation and Analysis

In Experiment 2, first information type, second information type, measurement time, and their interaction were included as fixed effects. We included a by-subject random intercept and a by-item random intercept for the combination of the pairs of cues. More complex models with random slopes failed to converge or resulted in singularity warnings.

Results

See Table 3 for descriptive statistics and Figure 2 for the results. We found a significant effect of measurement time, b = −0.49, SE = .03, $t(1001.32) = -17.33$, $p < .001$, meaning that impressions were significantly updated by the second piece of information provided. We also found a significant effect of first information type, b = −0.32, SE = .05, $t(74.85) = -6.61$, $p < .001$, which was qualified by a significant interaction with measurement time, b = −0.15, SE = .03, $t(1001.32) = -5.42$, $p < .001$. When comparing the updating effect on the two first information types, we found stronger updating for behavioral first impressions, b = −1.30, SE = .08, $t(1008.31) = -16.09$, $p < .001$, than for nonbehavioral impressions (either based on faces or voices), b = −0.68, SE = .08, $t(1008.31) = -8.42$, $p < .001$. Critically, we found a significant interaction effect of second information type and measurement time, b = 0.15, SE = .02, $t(1001.32) = 5.34$, p <.001: Impressions were updated significantly by both faces, b = −0.68, SE = .08, $t(1008.31) = -8.41$, $p < .001$, and voices, b = −1.29, SE = .08, $t(1008.31) = -16.16$, $p <$ .001, but the updating power of voices was significantly stronger than that of faces, regardless of the type of information in the first impression. There were no other significant effects, $b < -0.08$, SE = .05, $t < 1.74$, $p < .085$.

Discussion

Experiment 2 showed that both faces and voices can update impressions generated via behavioral and nonbehavioral information. Voices were more effective than faces in

Table 3. Experiment 2.

	Behavior						Nonbehavior					
	Time 1		Time 2		Score		Time 1		Time 2		Score	
Second information	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Face	−0.90	1.32	0.07	1.36	0.98	1.54	0.18	1.33	0.57	1.18	0.39	1.35
Voice	−0.98	1.28	0.64	1.29	1.62	1.53	−0.15	1.08	0.81	1.12	0.97	1.33

Note. Means and standard deviations of the evaluation given to the targets as a function of type of the first information (Behavior vs. Non-behavior), second information type (Face vs. Voice), and time (Time 1 vs. Time 2). The nonbehavioral information at Time 1 is always the other nonbehavioral cue (i.e., face) compared with that at Time 2 (i.e., voice). The column Score represents the difference between the evaluations given at the two stages (Time 2 – Time 1). Judgments are reported on the original scale ranging from −3 (negative) to 3 (positive).

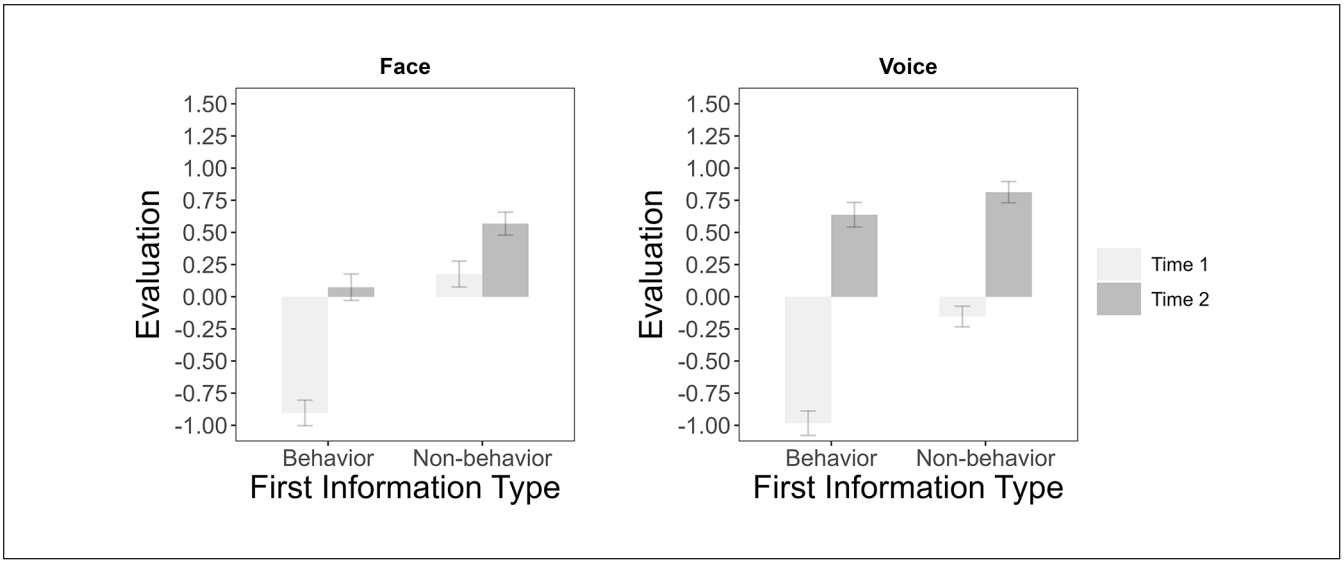

Figure 2. Bar Graph of Experiment 2 Results With 95% Confidence Intervals.
Note. The titles distinguish between the second information types (Face vs. Voice). First information type (Behavior vs. Non-behavior) is depicted on the x-axis. The nonbehavioral information at Time 1 is always the other nonbehavioral cue (e.g., face) compared with that at Time 2 (e.g., voice). The light gray bar represents the Time 1 evaluation, and the dark gray bar represents the Time 2 evaluation.

updating first impressions. Although such results were in harmony with Masi et al. (2022), they diverged from what was found in Experiments 1a-1b, where faces and voices were equally effective in updating behavior-based impressions. To address these mixed findings, we conducted a preregistered third experiment.

Experiment 3

Experiment 3 aimed at clarifying the mixed findings observed in Experiments 1a-1b and Experiment 2. One important difference between the two investigations was the experimental design. In Experiments 1a-1b, the type of second information (driving impression updating) was manipulated within participants, while in Experiment 2 it was manipulated between participants. One possibility is that the difference in the influence of voices and faces in updating impressions is confined to specific methodological conditions. Second, in Experiments 1a-1b, for each participant, faces and voices were directly compared with behaviors at Time 2. As behaviors were the most effective in updating impressions, one possibility is that comparing both faces and voices directly with behaviors may have induced participants to discard nonbehavioral cues as "marginal" or "less informative," ultimately reducing their difference. Moreover, given that attractiveness can impact impression formation from nonbehavioral cues and contribute to global evaluations, alongside trustworthiness (Lan et al., 2022; Rougier & De Houwer, 2024; Todorov, 2008), Experiment 3 ensured a more precise cue selection by employing a set of faces and voices that were perceived as both equally trustworthy and attractive. Thus, Experiment 3 tested the hypothesis that voices are superior to faces in updating impressions by employing a within-subject design that included only nonbehavioral (faces and voices) information

balanced on trustworthiness and attractiveness. We preregistered the study on OSF (https://osf.io/w38bx/).[3]

Method

Design. The experiment consisted of a 2 (Second information type: Face vs. Voice) × 2 (Measurement time: Time1 vs. Time2) within-subjects design.

Participants and Power Analysis. We preregistered a sample size determination analysis for a paired-sample *t* test with a target effect of $d_z = .25$ (small to medium effect, Funder & Ozer, 2019), with high power of 95%, at alpha = .05 (G*power, Faul et al., 2007), targeting the difference between faces and voices in the impression updating score obtained by subtracting the rating at Time 1 from that at Time 2, which is the effect we observe from the interaction between information type and time. The analysis returned a sample of 210 participants ($M_{age} = 30.49$, $SD_{age} = 8.74$, 112 females). As we employed a different method of analysis, we discussed the power of the critical effects in the context of a linear mixed effect model. Our safeguard sensitivity power analysis on the 20th percentile of the observed effect resulted in a sufficiently high power for detecting the updating effect of both faces and voices (power > 99%). We did not conduct a safeguard analysis on their difference in updating power because it was not significant.

Materials. Ten behaviors perceived as untrustworthy ($M = -1.15$, $SD = 0.28$) were taken from Experiment 2 stimuli and employed as first impression cues. For facial and vocal stimuli, we employed part of the cues used in Experiment 2 and re-pretested them for trustworthiness and attractiveness (see Supplementary Materials). The selected five faces and five voices were perceived as equally trustworthy ($M_{voices} = 0.69$, $SD_{voices} = 0.08$; $M_{faces} = 0.69$, $SD_{faces} = 0.17$) and did not differ in perceived attractiveness, on which they were evaluated as close to the neutral midpoint $M_{voices} = 0.07$, $SD_{voices} = 0.13$; $M_{faces} = 0.09$, $SD_{faces} = 0.511$).

Procedure. Participants were rewarded £8/hr, £0.40 for a 2.60-min experiment on Prolific Academic. The experiment implemented the same procedure as Experiment 1a except that information presented at Time 2 was either the target face or voice (not a behavior). Thus, the first piece of information consisted of a single negative behavior written below targets' names, presented one at a time and in random order. After expressing their judgment using the same global impression scale as in the previous experiments (Time 1), participants saw the positive face of one target and heard the voice of the other. Targets were presented in the same order as before, but the association of the target name with the new cue, either a face or a voice, was at random. Next, participants indicated their global impression of the two targets once again (Time 2). Finally, they were debriefed and thanked for their participation.

Table 4. Experiment 3.

	Time 1		Time 2		Score	
Second information	*M*	*SD*	*M*	*SD*	*M*	*SD*
Face	−1.10	1.31	0.05	1.21	1.15	1.52
Voice	−1.33	1.20	0.009	1.19	1.33	1.35

Note. Means and standard deviations of the evaluation given to the targets as a function of time (Time 1 vs. Time 2) and the second information type (Face vs. Voice). The column Score represents the difference between the evaluations given at the two stages (Time 2 – Time 1). Judgments are reported on the original scale ranging from −3 (negative) to 3 (positive).

Data Preparation and Analysis

We preregistered to analyze the data with a linear model (see Supplementary Materials for the results). However, to increase generalizability of the findings beyond the available cues and for consistency with the previous studies, we decided to deviate from the preregistration and include random effects as well, improving the quality of the results. Thus, second information type, measurement time, and their interaction were included as fixed effects. We also included a by-subject random slope accounting for differences between the two measurement times and a by-item random intercept for the combination of the pairs of cues. More complex models with random slopes failed to converge or resulted in singularity warnings.

Results

Table 4 provides descriptive statistics and Figure 3 the results. We found a main effect of measurement time, meaning that on average impressions were always significantly updated, $b = -0.62$, $SE = .04$, $t(209) = -15.61$, $p < .001$. The main effect of information type was not significant, $b = 0.06$, $SE = .05$, $t(68.89) = 1.16$, $p = .248$. Although the descriptive pattern was consistent with our hypothesis, suggesting that impressions tended to be updated more by voices than by faces, the interaction between the two factors was not significant, $b = 0.05$, $SE = .03$, $t(339) = 1.53$, $p = .126$.

Discussion

Experiment 3 tested the difference between faces and voices in a within-subjects design and without additional behavioral information. We found no significant difference between the updating of the two types of information. This experiment provides additional evidence that the difference between voices and faces in updating impressions based on behaviors may be influenced by multiple factors.

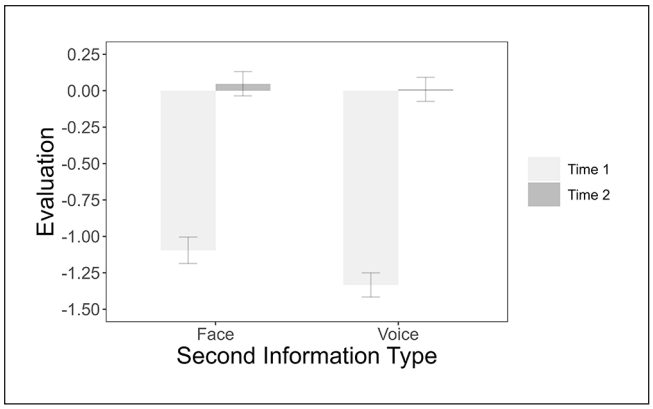

Figure 3. Bar Graph of Experiment 3 Results With 95% Confidence Intervals.
Note. Second information type (Face vs. Voice) is depicted on the x-axis, the light gray bar represents the Time 1 evaluation, and the dark gray bar represents the Time 2 evaluation.

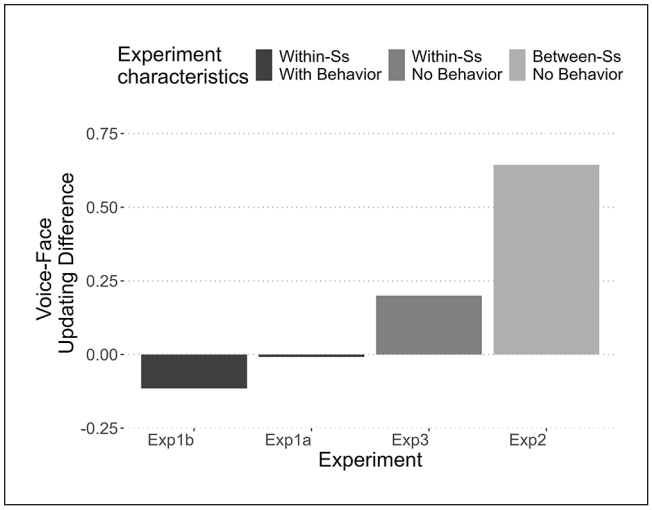

Figure 4. Summary of the Results of the Experiments.
Note. The y-axis represents the voice–face difference in updating. The x-axis represents the four experiments, ordered by size difference. The shades of gray represent the differences between the experiments as a combination of design (within-subjects or between-subjects) and the presence of the additional behavior along the face and the voice.

General Discussion

Impressions of individuals are often the results of conflicting information (Cone et al., 2017). Various cues contribute to forming and updating impressions, yet research on updating has primarily focused on behavior, rarely integrating other cues (but see Masi et al., 2022; Shen et al., 2020). Here, we investigated the efficacy of positive cues in updating negative first impressions, based on both behavioral and nonbehavioral information. In Experiments 1a-1b, participants formed negative impressions of targets displaying untrustworthy behavior. Subsequently, trustworthy information in the form of behaviors, faces, or voices was presented in a within-subjects design. We found that all cues could update behavior-based impressions, with behavioral information being the most effective. Faces and voices updated impressions comparably. In Experiment 2, a direct comparison between the updating impact of faces and voices to establish the relative superiority of each cue (this time manipulated between participants) showed that voices were more effective than faces in updating (both behavioral and nonbehavioral) impressions. Experiment 3 returned to a within-subject design while controlling for cue attractiveness; we found evidence for effective updating by both faces and voices, without significant evidence for the latter being stronger than the former.

This study represents a significant step forward in understanding impression updating (Cone et al., 2017; Masi et al., 2022; Shen et al., 2020). Our findings consistently show that facial images or vocal recordings can modify impressions based on behaviors. Importantly, they underscore the importance of nonbehavioral information, which continues to shape inferences despite the presence of potentially more diagnostic cues (i.e., behaviors). This aligns with previous research on the interplay between nonbehavioral information and behaviors in impression formation (e.g., Ko et al., 2009; Rezlescu et al., 2012). We extended such evidence to impression updating, revealing the enduring influence of nonbehavioral cues on impression dynamics. Therefore, these results fit well with an ecological perspective on social perception (McArthur & Baron, 1983), which suggests that people may be motivated to use multiple channels of information to form a comprehensive understanding of others. According to this theory, the evidence of impression updating from faces and voices may mean that they are considered informative cues that can influence participants' subsequent actions toward the target individual.

Our investigation also examined the differential influence of vocal and facial cues on the updating of behavioral impressions. The hypothesized difference between faces and voices was based on evidence from previous studies (Hansen et al., 2018; Masi et al., 2022) and the possibility that a voice might hold more powerful properties (e.g., information richness and agency; Kraus, 2017; Lavan et al., 2021). Such differences could make people assign more weight or informativeness to the social information (i.e., trustworthiness) conveyed by vocal cues rather than facial cues or, in other words, voices might be perceived as more diagnostic information. Our studies suggest that voices may be superior to faces although its greater diagnosticity depend on some contextual factors. Across experiments, we varied two features: the design (between-subjects vs. within-subjects) and the comparison with behavioral information when the face and voice were shown (Figure 4 summarizes the results of the face–voice comparison). A marked voice superiority was documented in Experiment 2, using a between-subjects design similar to that used in previous studies (Hansen et al., 2018; Masi et al., 2022). This design forces the perceiver to focus

on either the voice or the face without comparing the two cues within the same experiment. Thus, in conditions where only one cue, either the voice or the face, is available to update impressions, negative first impressions changed more due to positive voices than positive faces. This is consistent with previous findings in cross-modal impression updating (Masi et al., 2022). In a within-subjects design, i.e., when individuals were presented with both types of information, the difference was either null, as in Experiments 1a-1b (where participants were also exposed to behavioral information), or only descriptively in favor of the voice, as in Experiment 3.

We might speculate that people found voices more informative or diagnostic than faces for behavioral impression updating when presented alone (Experiment 2), but the direct comparison with alternative cue types, such as the face or a behavior, may have caused people to be more cautious in their judgments. The well-established contextual dependency of most social judgments may explain these results (Barker & Imhoff, 2021; Goller et al., 2018; Mussweiler, 2003). The comparison context may have shifted the judgment criterion when multiple cues were presented simultaneously rather than in isolation. It is possible that when the voice was presented with the face after a behavior-based impression had been formed (Experiment 3), they were assimilated, that is, the two cues were judged to be more similar than different. This could be due to their common nonbehavioral nature, which was compared with the type of the first information (i.e., behaviors) on which impressions were formed. Thus, their updating effects were not different. This phenomenon may have been even more pronounced when behavioral information is presented for comparison with a face or a voice at the second stage (i.e., nonbehavioral information, Experiments 1a-1b). In this case, the similarity between the first and second behavioral information may have made this latter a favorable type of information for updating while increasing the contrast between the behavioral and nonbehavioral information. Consequently, nonbehavioral information was assimilated and significantly reduced in updating power relative to the behavior (Mussweiler, 2003). Our findings emphasize the importance of considering the boundaries that influence the relative power of faces and voices on impression updating. Overall, this research complicates the framework described by previous research on impression updating, which has mostly used between-subjects designs (Cone et al., 2017; Masi et al., 2022; Shen et al., 2020; but see Brambilla et al., 2019), by suggesting that whether one cue leads to more updating than another also depends on the comparison context. Thus, this is something to take into account in future research.

The present work is not without limitations. Our findings should be contextualized within the impression updating scenario, particularly with respect to cues manipulated for trustworthiness. Similar to findings in other face–voice impression studies, the superiority of a cue may depend on the trait under consideration. Therefore, future studies should extend the investigation to other traits, such as sociability or dominance, where research on impression updating is scarce. At the same time, research should include the positive-to-negative pathway that was neglected here; testing traits characterized by positive asymmetry (e.g., competence), the pattern of results may differ from ours (Brambilla et al., 2021). Moreover, in our work, we used different types of face stimuli and pitch-manipulated voices to provide a variety of speaker choices. Despite rigorous pretesting to control for various influencing variables (e.g., attractiveness, national prototypicality differences, see Supplementary Materials), the stimuli represent only a few potential variations of faces and voices, which might affect the generalizability of the results. For example, we emphasize the absence of female faces and voices in our studies, whose vocal and facial characteristics as well as the stereotypes associated with these variations differ significantly from those of males (e.g., Ko et al., 2009). These factors serve as potential moderators and could alter the results.

This research provided one of the first evidence of the interconnectedness of behaviors, faces, and voices in impression updating. We showed that it takes only a few words or a single glance to update an impression based on behavior, and that vocal cues can be more powerful than facial cues, but only under certain circumstances. Taken together, our findings expand our knowledge of the malleability of social impressions.

Acknowledgments

We thank Prof. Marcello Gallucci for his comments on some of our statistical analyses.

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) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Matteo Masi https://orcid.org/0000-0001-6436-1663
Simone Mattavelli https://orcid.org/0000-0002-8934-8016

Data Availability Statement

All the materials, data, and analyses code are available at https://osf.io/de3rp/

Supplemental Material

Supplemental material is available online with this article.

Notes

1. Participants were exposed to the three targets and in the following experiments to two targets. This was consistent with

previous research (Masi et al., 2022; Shen et al., 2020) and helped to avoid fatigue and make sure participants paid attention to all the cues provided.

2. Because of the way variance is partitioned in linear mixed models, there is no agreed-upon way to calculate standardized effect sizes for individual model terms. Therefore, in this and the following experiments, we decided not to include a standardized effect size.
3. Prior to conducting Experiment 3, we ran another preregistered study (https://osf.io/98fam/) with the same goal that also considered differences in facial cues used in our previous experiments. However, a technical error occurred affecting the number of cues across conditions. Results of this study are reported in the Supplementary Materials.

References

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology*, *32*, 201–271. https://doi.org/10.1016/S0065-2601(00)80006-4

Asch, S. E., & Zukier, H. (1984). Thinking about persons. *Journal of Personality and Social Psychology*, *46*(6), 1230–1240. https://doi.org/10.1037/0022-3514.46.6.1230

Barker, P., & Imhoff, R. (2021). The dynamic interactive pattern of assimilation and contrast: Accounting for standard extremity in comparative evaluations. *Journal of Experimental Social Psychology*, *97*, 104190. https://doi.org/10.1016/j.jesp.2021.104190

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Brambilla, M., Carraro, L., Castelli, L., & Sacchi, S. (2019). Changing impressions: Moral character dominates impression updating. *Journal of Experimental Social Psychology*, *82*, 64–73. https://doi.org/10.1016/j.jesp.2019.01.003

Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. P. (2021). The primacy of morality in impression development: Theory, research, and future directions. *Advances in Experimental Social Psychology*, *64*, 187–262. https://doi.org/10.1016/bs.aesp.2021.03.001

Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology*, *108*(1), 37–57. https://doi.org/10.1037/pspa0000014

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. *Advances in Experimental Social Psychology*, *56*, 131–199. https://doi.org/10.1016/bs.aesp.2017.03.001

De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, *7*(10), 460–467. https://doi.org/10.1016/j.tics.2003.08.014

Fasoli, F., Maass, A., Paladino, M. P., & Sulpizio, S. (2017). Gay- and lesbian-sounding auditory cues elicit stereotyping and discrimination. *Archives of Sexual Behavior*, *46*, 1261–1277. https://doi.org/10.1007/s10508-017-0962-0

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fiske, S. T. (1992). Thinking is for doing: Portraits of social cognition from Daguerreotype to laserphoto. *Journal of Personality and Social Psychology*, *63*(6), 877–889. https://doi.org/10.1037/0022-3514.63.6.877

Foo, Y. Z., Sutherland, C. A. M., Burton, N. S., Nakagawa, S., & Rhodes, G. (2021). Accuracy in facial trustworthiness impressions: Kernel of truth or modern physiognomy? A meta-analysis. *Personality and Social Psychology Bulletin*, *48*, 1580–1596.

Freeman, J. B., Stolier, R. M., & Brooks, J. A. (2020). Dynamic interactive theory as a domain-general account of social perception. *Advances in Experimental Social Psychology*, *61*, 237–287. https://doi.org/10.1016/bs.aesp.2019.09.005

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, *14*(2), 214–237. https://doi.org/10.1177/1088868309359288

Goller, J., Leder, H., Cursiter, H., & Jenkins, R. (2018). Anchoring effects in facial attractiveness. *Perception*, *47*(10–11), 1043–1053. https://doi.org/10.1177/0301006618802696

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Hansen, K., Rakić, T., & Steffens, M. C. (2017). Competent and warm? *Experimental Psychology*, *64*(1), 27–36. https://doi.org/10.1027/1618-3169/a000348

Hansen, K., Rakić, T., & Steffens, M. C. (2018). Foreign-looking native-accented people: More competent when first seen rather than heard. *Social Psychological and Personality Science*, *9*(8), 1001–1009. https://doi.org/10.1177/1948550617732389

Jiang, Z., Li, D., Li, Z., Yang, Y., Liu, Y., Yue, X., . . . Xue, P. (2024). Comparison of face-based and voice-based first impressions in a Chinese sample. *British Journal of Psychology*, *115*(1), 20–39. https://doi.org/10.1111/bjop.12675

Johnson, K. L., Lick, D. J., & Carpinella, C. M. (2015). Emergent research in social vision: An integrated approach to the determinants and consequences of social categorization. *Social and Personality Psychology Compass*, *9*(1), 15–30. https://doi.org/10.1111/spc3.12147

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2698–2704. https://doi.org/10.1098/rspb.2012.0311

Ko, S. J., Judd, C. M., & Stapel, D. A. (2009). Stereotyping based on voice in the presence of individuating information: Vocal femininity affects perceived competence but not warmth.

Personality and Social Psychology Bulletin, *35*(2), 198–211. https://doi.org/10.1177/0146167208326477

Kraus, M. W. (2017). Voice-only communication enhances empathic accuracy. *American Psychologist*, *72*(7), 644–654. https://doi.org/10.1037/amp0000147

Kruse, F., & Degner, J. (2021). Spontaneous state inferences. *Journal of Personality and Social Psychology*, *121*(4), 774–791. https://doi.org/10.1037/pspa0000232

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), 33267. https://doi.org/10.1525/collabra.33267

Lan, M., Peng, M., Zhao, X., Chen, H., Liu, Y., & Yang, J. (2022). Facial attractiveness is more associated with individual warmth than with competence: Behavioral and neural evidence. *Social Neuroscience*, *17*(3), 225–235. https://doi.org/10.1080/17470919.2022.2069152

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, *21*(4), R143–R145. https://doi.org/10.1016/j.cub.2010.12.033

Lavan, N., Mileva, M., Burton, A. M., Young, A. W., & McGettigan, C. (2021). Trait evaluations of faces and voices: Comparing within- and between-person variability. *Journal of Experimental Psychology: General*, *150*(9), 1854–1869. https://doi.org/10.1037/xge0001019

Lenth, R. V. (2022). *Emmeans: Estimated marginal means, aka least-squares means* (R package, Version 1.8.1-1). https://CRAN.R-project.org/package=emmeans

Liu, M., Sommer, W., Yue, S., & Li, W. (2023). Dominance of face over voice in human attractiveness judgments: ERP evidence. *Psychophysiology*, *60*(11), Article e14358. https://doi.org/10.1111/psyp.14358

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Masi, M., & Fasoli, F. (2022). When fluency matters: The interplay between categorization fluency and gender atypicality on gaydar judgments. *Journal of Language and Social Psychology*, *41*(6), 746–759. https://doi.org/10.1177/0261927X221111382

Masi, M., Mattavelli, S., Fasoli, F., & Brambilla, M. (2022). Cross-modal impression updating: Dynamic impression updating from face to voice and the other way around. *British Journal of Social Psychology*, *61*(3), 808–825. https://doi.org/10.1111/bjso.12511

McAleer, P., & Belin, P. (2018). The perception of personality traits from voices. In S. Fruholz & P. Belin (Eds.), *The Oxford handbook of voice perception* (pp. 585–606). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.26

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'hello'? Personality impressions from brief novel voices. *PLoS ONE*, *9*(3), Article e90779. https://doi.org/10.1371/journal.pone.0090779

McArthur, L. Z., & Baron, R. M. (1983). Toward and ecological theory of social perception. *Psychological Review*, *90*(3), 215–238. https://doi.org/10.1037/0033-295X.90.3.215

McCarthy, R. J., & Skowronski, J. J. (2011). What will Phil do next? Spontaneously inferred traits influence predictions of behavior. *Journal of Experimental Social Psychology*, *47*(2), 321–332. https://doi.org/10.1016/j.jesp.2010.10.015

Mileva, M., & Lavan, N. (2023). Trait impressions from voices are formed rapidly within 400 ms of exposure. *Journal of Experimental Psychology. General*, *152*(6), 1539–1550. https://doi.org/10.1037/xge0001325

Mileva, M., Tompkinson, J., Watt, D., & Burton, A. M. (2018). Audiovisual integration in social evaluation. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 128–138. https://doi.org/10.1037/xhp0000439

Mileva, M., Tompkinson, J., Watt, D., & Burton, A. M. (2020). The role of face and voice cues in predicting the outcome of student representative elections. *Personality and Social Psychology Bulletin*, *46*(4), 617–625. https://doi.org/10.1177/0146167219867965

Moskowitz, G. B., Olcaysoy Okten, I., & Schneid, E. (2022). The updating of first impressions. In E. Balcetis & G. B. Moskowitz (Eds.), *The handbook of impression formation* (pp. 348–392). Routledge.

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, *110*(3), 472–489. https://doi.org/10.1037/0033-295X.110.3.472

Newman, L. S. (1996). Trait impressions as heuristics for predicting future behavior. *Personality and Social Psychology Bulletin*, *22*(4), 395–411. https://doi.org/10.1177/0146167296224006

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, *34*(2), 83–110. https://doi.org/10.1007/s10919-009-0082-1

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science*, *9*(3), 319–332. https://doi.org/10.1177/1745691614528519

Rakić, T., Steffens, M. C., & Mummendey, A. (2011). When it matters how you pronounce it: The influence of regional accents on job interview outcome. *British Journal of Psychology*, *102*(4), 868–883. https://doi.org/10.1111/j.2044-8295.2011.02051.x

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PLOS ONE*, *7*(3), Article e34293. https://doi.org/10.1371/journal.pone.0034293

Rezlescu, C., Penton, T., Walsh, V., Tsujimura, H., Scott, S. K., & Banissy, M. J. (2015). Dominant voices and attractive faces: The contribution of visual and auditory information to integrated person impressions. *Journal of Nonverbal Behavior*, *39*(4), 355–370. https://doi.org/10.1007/s10919-015-0214-8

Rougier, M., & De Houwer, J. (2024). Updating stereotypical attributions in light of new information: The attractiveness halo effect changes when attractiveness changes. *European Journal of Social Psychology*, *54*(1), 364–379. https://doi.org/10.1002/ejsp.3017

Rule, N. O., & Sutherland, S. L. (2017). Social categorization from faces: Evidence from obvious and ambiguous groups. *Current Directions in Psychological Science*, *26*(3), 231–236. https://doi.org/10.1177/0963721417697970

Rule, N. O., Tskhay, K. O., Brambilla, M., Riva, P., Andrzejewski, S. A., & Krendl, A. C. (2015). The relationship between antigay prejudice and the categorization of sexual orientation. *Personality and Individual Differences*, *77*, 74–80.

Schneid, E. D., Crawford, M. T., Skowronski, J. J., Irwin, L. M., & Carlston, D. E. (2015). Thinking about other people: Spontaneous trait inferences and spontaneous evaluations. *Social Psychology*, *46*(1), 24–35. https://doi.org/10.1027/1864-9335/a000218

Shen, X., Maiolatesi, A. J., Mann, T. C., & Ferguson, M. J. (2023). The Relation between updated implicit evaluations and the Trust Game. *Social Psychological and Personality Science*, *14*(8), 911–921. https://doi.org/10.1177/19485506221141661

Shen, X., Mann, T. C., & Ferguson, M. J. (2020). Beware a dishonest face? Updating face-based implicit impressions using diagnostic behavioral information. *Journal of Experimental Social Psychology*, *86*, 103888. https://doi.org/10.1016/j.jesp.2019.103888

Skowronski, J. J., Carlston, D. E., & Hartnett, J. (2008). Spontaneous impressions derived from observations of behavior: What a long, strange trip it's been (and it's not over yet). In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 313–333). Guilford Publications.

Snyder, M., & Stukas, A. A., Jr. (1999). Interpersonal processes: The interplay of cognitive, motivational, and behavioral activities in social interaction. *Annual Review of Psychology*, *50*, 273–303. https://doi.org/10.1146/annurev.psych.50.1.273

Sutherland, C. A., & Young, A. W. (2022). Understanding trait impressions from faces. *British Journal of Psychology*, *113*(4), 1056–1078. https://doi.org/10.1111/bjop.12583

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, *1124*(1), 208–224. https://doi.org/10.1196/annals.1440.012

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*(5), 1051–1065. https://doi.org/10.1037/0022-3514.83.5.1051

Uleman, J. S., & Kressel, L. M. (2013). A brief history of theory and research on impression formation. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition* (pp. 53–73). Oxford University Press.

Vives, M. L., Frances, C., & Baus, C. (2023). Automating weighing of faces and voices based on cue saliency in trustworthiness impressions. *Scientific Reports*, *13*(1), 20037.

Vonk, R. (1994). Trait inferences, impression formation, and person memory: Strategies in processing inconsistent information about persons. *European Review of Social Psychology*, *5*(1), 111–149.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and voice perception: Understanding commonalities and differences. *Trends in Cognitive Sciences*, *24*(5), 398–410. https://doi.org/10.1016/j.tics.2020.02.001

Zebrowitz, L. A. (2017). First impressions from faces. *Current Directions in Psychological Science*, *26*(3), 237–242. https://doi.org/10.1177/0963721416683996

Zuckerman, M., & Sinicropi, V. (2011). When physical and vocal attractiveness differ: Effects on favorability of interpersonal impressions. *Journal of Nonverbal Behavior*, *35*(2), 75–86. https://doi.org/10.1007/s10919-011-0106-5