

# Theoretical and legal issues related to choice blindness for voices

Citation for published version (APA):

Sauerland, M., Sagana, A., & Otgaar, H. (2013). Theoretical and legal issues related to choice blindness for voices. *Legal and Criminological Psychology*, 18(2), 371-381. <https://doi.org/10.1111/j.2044-8333.2012.02049.x>

## Document status and date:

Published: 01/09/2013

## DOI:

[10.1111/j.2044-8333.2012.02049.x](https://doi.org/10.1111/j.2044-8333.2012.02049.x)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.



## Theoretical and legal issues related to choice blindness for voices

Melanie Sauerland\*, Anna Sagana and Henry Otgaar

Department of Clinical Psychological Science, Maastricht University,  
The Netherlands

**Purpose.** To examine whether choice blindness occurs for auditory stimuli, namely voices.

**Methods.** One hundred participants listened to three pairs of voices and had to decide each time which one they found more sympathetic or sounded more criminal. After they made a choice, participants were presented with the chosen voice again and had to match it to a face. However, during the second trial, participants were actually presented with the voice they had previously *not* chosen.

**Results.** Only 19% of the participants detected this change concurrently, an additional 10% detected it retrospectively. This indicates that choice blindness transfers to auditory stimuli. Whether participants had previously evaluated sympathy or criminality of the voices had no effect on choice blindness.

**Conclusions.** The study shows that choice blindness is a robust phenomenon that can also be elicited when auditory stimuli are employed. Implications for earwitness testimony and expert witnesses in the context of forensic speech analysis are discussed.

While most of us would probably agree that we are free to make choices every day and that we are aware of the choices we make, a series of experiments recently challenged this assumption. That is, Johansson, Hall, Sikström, and Olsson (2005) instructed 120 participants to decide which of two female faces they found more attractive. After they made a decision, participants were handed the chosen face and were asked to motivate their decision. In 3 of the 15 trials, however, the participants' choice was manipulated. Here, a magical card trick was used in which participants ended up with the very face they had not chosen. Interestingly, only 13% of the manipulated trials were detected at the time of the manipulation. The authors dubbed this phenomenon *choice blindness*.

Also, another significant result from this study was that some participants included facial features in their choice motivation that the (manipulated) face in front of them did not carry (e.g., 'she had a nice smile' while the person on the photo was not smiling).

---

\*Correspondence should be addressed to Melanie Sauerland, Section Forensic Psychology, Faculty of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD Maastricht, the Netherlands (e-mail: melanie.sauerland@maastrichtuniversity.nl).

Others gave reasons that could only refer to the manipulated choice (e.g., refer to a smile when only the manipulated photo displayed a smile). Furthermore, during a post-test interview, 84% of the participants who failed to detect any of the manipulations stated that they thought they would be able to do so. This effect has been named *choice blindness blindness* (Johansson *et al.*, 2005).

Since this first study on choice blindness, the effect has been demonstrated for a wide variety of visual stimuli such as female faces and abstract patterns (Johansson, Hall, & Sikström, 2008), as well as olfactory (smell of tea) and gustatory (taste of jams) stimuli (Hall, Johansson, Tärning, Sikström, & Deutgen, 2010). Furthermore, choice blindness seems to be a robust phenomenon that persists across different stimulus similarities (although the effect tends to be stronger for more similar stimuli) and stimulus exposure times (e.g., 2 s vs. 5 s vs. unlimited; Johansson *et al.*, 2005). Additionally, choice blindness is not limited to live administration as a real-world magic trick, but has also been found on a computerized test (Johansson *et al.*, 2008).

To further test the robustness of the choice blindness effect, the current study examined whether this phenomenon would also be present when using auditory stimuli, specifically voices. We expected to find the effect for two reasons. First, the effect has been demonstrated for three different modalities so far (Hall *et al.*, 2010). We see no indication why auditory stimuli should be an exception. The second argument focuses more closely on the performance evolving around face and voice perception. Although it is generally found that people are better at recognizing faces than at recognizing voices (e.g., Hanley & Damjanovic, 2009; Olsson, 2000; Stevenage, Howland, & Tippelt, 2011), there are some parallels in the research findings looking at voices and faces. While humans are quite good at recognizing familiar voices, recognizing unfamiliar ones is much more difficult (e.g., Yonan & Sommers, 2000). The same holds for recognizing familiar versus unfamiliar faces (Johnston & Edmonds, 2009). On the other hand, humans' ability to *discriminate* voices is high – although not perfect (e.g., Kreiman & Papcun, 1991), as it is for discriminating faces (Jenkins, White, Van Montfort, & Burton, 2011). In the light of these parallels and given the repeated finding of choice blindness for faces, we reason that the effect should occur for voices as well. If our hypothesis is supported, this could have important practical implications, namely for evidence given by earwitnesses and by expert witnesses in the context of forensic speech analyses. We will return to this in the discussion.

## Methods

### Participants

A total of  $N = 100$  (50 male, 50 female;  $M_{\text{age}} = 22.8$  years,  $SD_{\text{age}} = 1.8$ , age range: 20–30) native German speakers from Aachen and Maastricht University participated in the study. Participants were bachelor (87%) and master (12%) students with different majors. The participation was on voluntary basis; no monetary awards or course credits were provided. The study was approved by the standing ethical board of Maastricht University.

### Materials

#### Voice pairs

As stimulus material, we used three pairs of voices. The speakers, who were native male Germans, were instructed to read out loud, in normal tone and pace, three sentences

(in German). These sentences were taken from the train magazine of the German Railways.

In every ICE and IC there are coaches with express reservation or Bahn Comfort seats. Not all of these are always reserved. In ICEs of the first generation, these seats are located in the coaches with the numbers 6 and 7.

The voices were recorded with a high-quality sound recorder (BOSS Digital Recorder Micro BR) and stored in an MP3 format. The audio track duration ranged from 13 to 17 s ( $M = 14.5$  s) with an audio bit rate of 128 kbps.

Three pairs of voices varying in similarity (high, moderate, or low) were created. Voice similarity was established in a pilot study with  $N = 25$  native German speakers as participants who were second-year psychology students and received course credit for their participation. Specifically, participants listened to 21 pairs of voices and rated them on a 10-point Likert scale (1 = *very different*; 10 = *very similar*). Those voice pairs that were rated to be most similar ( $M = 6.19$ ,  $SD = 1.29$ ) or least similar ( $M = 4.00$ ,  $SD = 1.32$ ) were selected, as well as one voice pair of moderate similarity ( $M = 5.19$ ,  $SD = 1.63$ ). These three pairs of voices differed significantly in similarity,  $t_s(25) \geq 3.05$ ,  $p_s \leq .005$ .

#### Photos

Six male head and shoulders photos were selected from the local database of Maastricht University. The size of the photos was 12.5 cm  $\times$  16.0 cm and they were presented to the participants as printouts.

#### Post-test questionnaire

The post-test questionnaire was an adjusted version of the one used by Johansson *et al.* (2008) to match our stimuli. Furthermore, we added one more question. Specifically, participants were first asked whether they had noticed anything strange during the experiment, then whether they had noticed anything odd with the voices. If at this point no signs of detection were revealed, participants were then told that we planned a follow-up study where the choices sometimes would be switched and were asked if they would notice such a change. The answer to this question is referred to as the degree of choice blindness in 'Results' section. Finally, participants were informed that manipulations had actually taken place in the current study and were asked if they had noticed this. If participants responded with 'yes', they were also asked in which trial they had noticed the change (additional item compared to Johansson *et al.*, 2008).

#### Marlowe–Crowne Social Desirability Scale (MCSDS)

The MCSDS (Crowne & Marlowe, 1960) is a 33-item true/false self-report inventory that aims to assess the individual's need for approval. The scores range from 0 to 33, indicating a person's number of true responses, with a mean score of 13.71 ( $SD = 5.78$ ) for normal adults (Crowne & Marlowe, 1960). Examples of the items are: 'I like to gossip at times' or 'I never resent being asked a favor'. The rationale of the MCSDS is that persons with higher need for approval tend to give more socially desirable responses than the average (Leite & Beretvas, 2005). Here, we administered the MCSDS to examine whether choice blindness is associated with the tendency to act according to social demands, meaning

that people might fail to report that they did notice the change in the manipulated trial due to their need for approval.

#### *Gudjonsson Compliance Scale (GCS)*

The GCS (Gudjonsson, 1989) is a self-report inventory consisting of 20 true/false statements such as: 'People with a lot of authority make me feel uncomfortable', 'I give in easily when I am pressured', or 'I try to please others'. It is used to measure the tendency of people to conform to other people's requests. The scores range from 0 to 20. An average score of 8.9 ( $SD = 3.2$ ) is considered to be typical for normal adults (Gudjonsson & Sigurdsson, 2004). Similar to the MCSDS, the GCS was included to control for the possibility that participants were not blind to the manipulations, but rather complied with the presumed expectations of the experimenters.

#### **Design**

Similarity between the voices in the manipulated trial (high vs. low) was varied in a two-level one factorial between-subjects design. The detection rate of manipulated trials was the dependent variable. Two different measures of detection were used, namely concurrent and retrospective. Concurrent detection rate includes all participants who immediately noticed the manipulation when it had taken place. Retrospective detection additionally includes all those participants who reported that they had noticed the change of the voices in the post-test questionnaire.

#### **Procedure**

Data collection took place at two university libraries. Participants were approached and asked if they would like to participate in a study conducted by our University. All participants were tested individually. Upon agreement, participants signed the informed consent form and were randomly assigned to the conditions.

Then, participants received a written explanation of what the purpose of the study was (cover story). Some participants were led to believe that the aim of the study was to investigate whether features of a voice were associated with perceived sympathy in facial expressions. Other participants were informed that the university was involved in a research project that investigated whether auditory information could be beneficial for identifying criminals. Originally, this variation was meant to manipulate the perceived importance of the study as a second factor. However, a manipulation check showed that there was no difference in how important participants perceived the two conditions,  $p = .593$ . Therefore, we will not discuss this manipulation any further. Note that the two conditions also did not vary with regard to the choice blindness measures,  $ps = 1.000$ .

After reading the cover story, participants listened to the first pair of voices. The audio stimuli were presented via a laptop. To minimize external sound distraction, high-quality headphones that covered the ears completely were used. Participants then rated each of two voices regarding their sympathy or criminality on a 10-point Likert scale (1 = *not at all criminal/sympathetic*; 10 = *highly criminal/sympathetic*). Hereafter, participants had to decide which voice they ultimately found more sympathetic or criminal by marking a checkbox. The chosen voice was then directly played once again. In accordance with previous studies (Hall *et al.*, 2010; Johansson *et al.*, 2005), we used the same clips on both occasions. Now, participants were given the task to

match the voice to one of two photos that were presented in front of them. This was to ensure that participants would also listen closely the second time that they heard the voice. After participants indicated which picture matched the voice, the same procedure was followed for two more pairs of voices. On the second trial, however, the participants' choice was manipulated. Specifically, participants were presented with the voice they had *not* previously chosen. If the participant noticed the manipulation concurrently, the experimenter pretended having switched the voices by mistake and continued with the third trial. The order of presentation of the three voice pairs was moderate-high-low similarity for the high-similarity condition and moderate-low-high for the low-similarity condition. There was no condition in which the moderately similar voices were manipulated.

After the three trials were completed, the post-test questionnaire was conducted in order to assess whether the participants had detected the manipulation, but had for some reason not mentioned that. Hereafter, participants filled in the MCSDS and the GCS. Finally, participants were thanked and debriefed.

## Results

An *alpha* level of .05 was used for all inferential analyses. For non-parametric analyses of  $2 \times 2$  contingency tables, *phi* is reported as a measure of effect size.

### Choice blindness

The overall concurrent detection rate was 19% ( $n = 19$ ), with an additional 10% ( $n = 10$ ) of trials being detected retrospectively. Thus, 29% ( $n = 29$ ) of the participants detected the manipulated trial in total. Of all  $N = 100$  participants, 63 perceived themselves as able to detect the manipulation and of the  $n = 71$  participants who did not detect the manipulation, 38 (53.5%) thought they would be able to do so. Thus, considerable choice blindness was observed.

Two chi-square tests with voice similarity (high vs. low) as independent variable and concurrent vs. retrospective detection as dependent variables were conducted. The results revealed a significant effect of similarity on concurrent detection,  $\chi^2(1, N = 100) = 5.26, p = .040, \phi = 0.23$ . As expected, concurrent detection was more likely to arise for less similar voices ( $M = 28.0\%$ ), than for more similar ones ( $M = 10.0\%$ ). However, for retrospective detection, there was no significant effect of similarity,  $\chi^2(1, N = 100) = 2.38, p = .186, \phi = 0.15$ . Nevertheless, on a descriptive level, manipulations were detected more often for less similar ( $M = 36.0\%$ ) than for more similar voices ( $M = 22.0\%$ ).

### Choosing patterns

In the low-similarity condition, 18 participants chose the voice that was played first and 32 the voice that was played second. Similarly, in the high-similarity condition, 17 participants chose the voice that was played first and 33 the voice that was played second. This means that regardless of similarity condition, participants decided for the second voice in about two-thirds of the cases, whereas the voice played first was chosen in only about one-third of the cases. More interestingly, when participants chose the voice they heard first, they were less likely to discover the change than if they chose

the voice presented second. This was true for both concurrent detection,  $\chi^2(1, N = 100) = 6.18, p = .015, \phi = 0.25$ , and retrospective detection,  $\chi^2(1, N = 100) = 10.91, p = .001, \phi = 0.33$ . Although no separate tests could be computed for the two similarity conditions due the small *ns* in the individual cells, similar result patterns emerged for both conditions. Specifically, if participants chose voice 1 and were in the low-similarity condition, they spontaneously detected the change in only 2 of 18 cases (retrospective: 2 out of 16); for the high-similarity condition, it was 0 of 17 cases (retrospective: 1 out of 17 cases). If participants chose voice 2 in the low-similarity condition, they spontaneously detected the change in 5 of 33 cases (retrospective: 10 out of 33); for the high-similarity condition, it was 12 of 32 cases (retrospective: 16 out of 32 cases). These results may be due to recency effects and point to the role of short-term memory. Furthermore, the possible effects of distinctiveness will be explored at the end of the results section.

### **Compliance and social desirability measures**

It could be argued that some participants' did detect the change during the manipulated trial, but did not say so as a result of a tendency to respond in a socially desirable way or to comply. To control for such effects, participants MCSDS and GCS scores were obtained. Choice blind participants did not have higher social desirability scores ( $M = 14.3, 95\% \text{ CI} = 13.2\text{--}15.5$ ) or compliance scores ( $M = 7.1, 95\% \text{ CI} = 6.2\text{--}8.0$ ) than non-choice blind participants ( $M = 12.6, 95\% \text{ CI} = 12.4\text{--}16.4; M = 9.0, 95\% \text{ CI} = 6.9\text{--}9.2$ , respectively), all  $|t|s \leq 1.12, ps \geq .234, |d|s \leq 0.27$ . These results contradict the idea that choice blindness can be attributed to social demands or compliance.

### **Post hoc voice ratings**

To explore possible explanations for the effects found,  $N = 18$  native German speakers with different occupations ( $M_{\text{age}} = 38.3$  years,  $SD_{\text{age}} = 11.3$ ) rated the used voices regarding their attractiveness, distinctiveness, pitch, and presence of an accent. The wording of the items and the referring anchors of the employed 10-point Likert scales ranging from 1 to 10 can be found in Table 1. None of our three voice pairs differed with regard to perceived voice pitch or presence of an accent/dialect,  $|t|s(17) \leq 1.50, ps \geq .152, |d|s = \leq 0.36$ . Furthermore, the presence of an accent was generally estimated as being small ( $M = 2.57, SD = 1.78$ ). Note that all speakers originated from the same region

**Table 1.** Items for measuring attractiveness, distinctiveness, pitch of voice, and presence of accent/dialect of the selected voice pairs

	Item	Anchor low (1)	Anchor high (10)
Attractiveness	How attractive is this voice to you?	Not attractive at all	Very attractive
Distinctiveness	How easy would you find it to recognize this voice from a group of people who are talking to each other?	Very difficult	Very easy
Pitch of voice	How is the pitch of this voice?	Very low	Very high
Accent/dialect	Does this person speak with an accent/dialect?	No accent/dialect at all	Strong accent/dialect

in Western Germany. The two high-similarity voices differed somewhat with regard to their attractiveness ( $M_s = 5.61$  vs.  $4.28$ ),  $t(17) = 2.15$ ,  $p = .046$ ,  $d = 0.51$ , whereas the two low-similarity voices differed with regard to their distinctiveness ( $M_s = 4.06$  vs.  $5.50$ ),  $t(17) = -2.43$ ,  $p = .027$ ,  $d = -0.58$ . All other comparisons were non-significant.

## Discussion

The aim of the current study was to replicate the choice blindness effect for auditory stimuli, voice fragments in particular. To ensure maximum comparability, we kept the design as parallel to previous choice blindness experiments as possible. As expected, a large proportion of participants (71%) were blind towards the manipulation that occurred during one of the three trials. The observed 19% concurrent detection lies in the range of the 12–39% that have been found in facial choice blindness conditions (Johansson *et al.*, 2005, 2008). It is also comparable to the choice blindness rates found for olfactory and gustatory stimuli (Hall *et al.*, 2010). Thus, our study replicates and extends previous work on choice blindness thereby showing that choice blindness is a robust phenomenon that seems to operate in a comparable way for different senses.

While auditory stimuli seem to be affected by the choice blindness effect as much as other senses, our data also allow one tentative conclusion for a possible boundary condition of the effect. Specifically, the similarity between the chosen versus manipulated stimuli seems to be a limiting factor. We found an effect of similarity for concurrent (but not retrospective) detection and previous authors reported such an effect for a limited selection of stimuli (Hall *et al.*, 2010). Others, however, have not found the effect (Johansson *et al.*, 2005). The degree of similarity or lack thereof is most likely the determining factor. The *post hoc* voice ratings, however, showed that our low-similarity voices did not only differ in similarity, but also in distinctiveness, whereas our high similarity voices did not. This points to the possibility that differences in distinctiveness, possibly in addition to differences in similarity, may create a boundary condition for the effect. Previous choice blindness studies also have controlled for stimulus similarity, but not distinctiveness (Hall *et al.*, 2010; Johansson *et al.*, 2006, 2008). Therefore, it is conceivable that the effects found (or not found) there also go back to differences (or lack thereof) in distinctiveness. This might be an interesting issue to explore in future studies.

One argument that comes to mind when encountering choice blindness studies is that participants probably did notice the change but refrained from saying so. This could be because they were not confident enough to indicate the error or because they thought that declaring it during the experiment could spoil the results. In both cases, though, the straightforward thing to do during the post-test questionnaire would be to make mention of the recognition of the error then. Indeed, the demand characteristics during the post-test questionnaire should work against the choice blindness effect, not for it (Hall *et al.*, 2010). Furthermore, the genuine surprise displayed by many participants after the debriefing is hard to accommodate with the idea that participants pretended *not* to have noticed the change (cf. Hall *et al.*, 2010). Finally, the social desirability and compliance results speak against the interpretation that persons who are prone to the social demands of the situation are more prone to choice blindness. This is in agreement with Merckelbach, Jelicic, and Pieters (2011) who found that 49 participants (63%) were blind towards their own ratings given on two items of the Symptom Checklist-90 (Derogatis, Lipman, & Covi, 1973), independent of social desirability scores.

What other factors could be capable of explaining the choice blindness effect as found in the current study? Possible candidates could be attention and short-term memory. One specific pattern in our results suggests a prominent role of short-term memory. In total, participants were presented with three auditory fragments: voice 1 (fragment 1), voice 2 (fragment 2), and then, depending their choice again with voice 1 or 2 (fragment 3). Interestingly, participants who chose voice 1 in the crucial second trial were less likely to detect the change than participants who chose voice 2, regardless of voice similarity or distinctiveness. In cases where the participant chose voice 2, this voice should still have been in the phonological store of the phonological loop (Baddeley, 1992) when fragment 3 was played. However, if a participant chose voice 1, this voice should already have been absent from phonological store due to time constraints of this store and due to the presentation of fragment 2 prior to the presentation of fragment 3. Accordingly, it is reasonable to assume that it was more difficult for participants to make comparisons of fragments 1 and 3 than of fragments 2 and 3. On the other hand, one may argue that participants who chose fragment 1 should have noticed that fragment 3 was identical to the preceding fragment 2. This is where the role of attention may come to play. Participants may not have directed enough attention to the comparison between fragments 2 and 3 by the time they articulated their choice of fragment 1 and while they were busy matching fragment 3 to one of two faces. Thus, an interplay between short-term memory and attention could be responsible for the result pattern as found.

An alternative explanation could be that distinctiveness is the crucial player in the current findings. Distinctiveness has shown to be an important factor in face-recognition performance (Bruce, Burton, & Dench, 1994; Shapiro & Penrod, 1986; Valentine, 1991) as well as voice-recognition performance (Mullennix, Ross, Kuykendall, Conard, & Barb, 2011; Orchard & Yarmey, 1995). In our design, the voices of one voice pair were always presented in the same order. For the low-similarity condition, this means that the less-distinctive voice was always presented first. Thus, it is conceivable that the finding that detection was better when the second voice was chosen is a distinctiveness effect rather than a recency effect. Nevertheless, the two voices in the high-similarity condition did not differ in distinctiveness. Yet, the result patterns of the two conditions show great analogies. If distinctiveness played a superior role in the choice blindness effect, we would have expected the results patterns of the two conditions to deviate more from each other. Nevertheless, we cannot exclude the possibility that distinctiveness plays a viable role in the choice blindness effect. Alternatively, recency effects and distinctiveness may add up. This will be for future studies to determine.

Although the two high-similarity voices differed in perceived attractiveness, we do not think this to be a major factor for the current results, because at least facial attractiveness has repeatedly been shown to be unrelated to memory performance (Sarno & Alley, 1997; Wickham & Morris, 2003). We are not aware of any studies that tested the effect for voices, but given other parallels in the performance revolving around the perception of faces and voices, we suspect that voice attractiveness does not have an impact on performance. This is, however, subject to empirical testing.

We will now turn to possible implications of our findings for evidence given by earwitnesses and by expert witnesses in the context of forensic speech analyses. Cases exist in which a perpetrator's voice may be the only evidence available. This can concern any crime that occurs in the dark, via telephone, or that involves a masked perpetrator. In such cases, auditory identification testimony can be treated as direct evidence of identity and therefore can be presented in court (Clifford, 1980). After identifying a suspect from an auditory lineup, an earwitness may be required to identify him/her again inside

the courtroom. Consider a situation where the lineup administrator – intentionally or unintentionally – writes down a different choice than the one made by the witness. Given the current results, one must worry that a considerable proportion of earwitnesses may not notice that change. This, of course, could have devastating consequences for the defendant. Because such effects have previously been demonstrated for facial lineups (Sagana, Sauerland, & Merckelbach, 2012), it seems very likely that they would also occur for voice lineups.<sup>1</sup>

Let us take a closer look at the literature comparing face and voice recognition. In a series of studies, Hanley and colleagues have demonstrated that it is more difficult for people to retrieve semantic information about celebrities from a voice than from a face (Hanley, Smith, & Hadfield, 1998), even when the face has been blurred (Damjanovic & Hanley, 2007; Hanley & Damjanovic, 2009). To control for the fact that people encounter celebrities' faces more frequently than their voices (i.e., in the media), a recent study replicated the effect when using personally familiar people's voices and faces (Brédart, Barsics, & Hanley, 2009). These results have been interpreted such that the face-recognition system is more closely connected to biographical information stored in semantic memory than the voice-recognition system (Damjanovic & Hanley, 2007). Another interpretation could be that people are better at distinguishing between faces than between voices (Brédart *et al.*, 2009). In any case, both interpretations indicate that choice blindness for lineups as described above should actually be more pronounced for voices than for faces. This seems an important and interesting research question worth further exploration.

Our results may also be meaningful for forensic speech expert witnesses. These professionals often deal with different speech samples and have to decide whether two samples belong to the same person or not, thus linking suspects to certain crimes. This could involve speech samples collected in suspects' houses (e.g., answering machines, telephones), recorded by the investigative teams (e.g., in cases of bomb threats), obtained in wiretap operations as well as obtained directly from existing suspects. Forensic speech analysts can provide the police with important leads in a case and their testimony can eventually play a vital role during trial. Although the paradigm used in the current study deviates somewhat from the forensic speech analysis setting, our results might apply nevertheless. What if speech analysts are choice blind to auditory stimuli, just like the participants in our study? This would obviously question the validity of the evidence given by such experts. Although speech identification experts perform significantly better at identifying speakers than non-experienced persons (Schiller & Köster, 1998), our speculative guess is that the effect would not disappear completely. Further, these experts are unlikely to be questioned due to their expert status. Therefore, such a bias is likely to go unnoticed and execute a devastating impact on innocent suspects and the legal system as a whole. Future research should address these issues.

To conclude, the current study has successfully shown that the choice blindness effect is present for auditory stimuli. In our view, its application to the legal system opens a new interesting and important field that can inform us on how to avoid biases in legal procedures.

---

<sup>1</sup> The issue may be raised that we tested for clip recognition rather than voice recognition since we used the same clips in both phases of the Experiment (choosing and pairing with face). This is not a problem for the point we make here, however. In fact, the considerable choice blindness effect we found while using the same clips speaks to the prominence of the effect and we would expect the effect to be even stronger if different voice clips were used.

## Acknowledgements

The authors would like to thank Eva Billen, Joke de Boer, Lilith Boettcher, Aleksandra Litvinova, Lisa Müller-Ehrenberg, Nina Schmitz, Hille Springsfeld, Dajana Zelic, and Tanya Sauerland for their help in collecting data.

## References

- Baddeley, A. (1992). Working memory. *Science*, *255*, 556–559. doi:10.1126/science.1736359
- Brédart, S., Barsics, C., & Hanley, R. (2009). Recalling semantic information about personally known faces and voices. *European Journal of Cognitive Psychology*, *21*, 1013–1021. doi:10.1080/09541440802591821
- Bruce, V., Burton, A. M., & Dench, N. (1994). What's distinctive about a distinctive face? *The Quarterly Journal of Experimental Psychology*, *47A*, 119–141. doi:10.1080/14640749408401146
- Clifford, B. R. (1980). Voice identification by human listeners: On earwitness reliability. *Law and Human Behavior*, *4*, 373–394. doi:10.1007/BF01040628
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349–354. doi:10.1037/h0047358
- Damjanovic, L., & Hanley, J. R. (2007). Recalling episodic and semantic information about famous faces and voices. *Memory & Cognition*, *35*, 1205–1210. doi:10.1016/j.concog.2010.06.002
- Derogatis, L. R., Lipman, R. S., & Covi, L. (1973). SCL-90: An outpatient psychiatric rating scale—preliminary report. *Psychopharmacology Bulletin*, *9*, 13–28.
- Gudjonsson, G. H. (1989). Compliance in an interrogative situation: A new scale. *Personality and Individual Differences*, *10*, 535–540. doi:10.1016/0191-8869(89)90035-4
- Gudjonsson, G. H., & Sigurdsson, J. F. (2004). The relationship of suggestibility and compliance with self-deception and other-deception. *Psychology, Crime & Law*, *10*, 447–453. doi:10.1080/10683160310001634278
- Hall, L., Johansson, P., Tärning, B., Sikström, S., & Deutgen, T. (2010). Magic at the marketplace: Choice blindness for the taste of jam and the smell of tea. *Cognition*, *117*, 54–61. doi:10.1016/j.cognition.2010.06.010
- Hanley, J. R., & Damjanovic, L. (2009). It is more difficult to retrieve a familiar person's name and occupation from their voice than from their blurred face. *Memory*, *17*, 830–839. doi:10.1080/09658210903264175
- Hanley, J. R., Smith, T., & Hadfield, J. (1998). I recognise you but I can't place you. An investigation of familiar-only experiences during tests of voice and face recognition. *Quarterly Journal of Experimental Psychology*, *51A*, 179–195. doi:10.1080/027249898391819
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*, 313–323. doi:10.1016/j.cognition.2011.08.001
- Johansson, P., Hall, L., & Sikström, S. (2008). From change blindness to choice blindness. *Psychologia: An International Journal of Psychology in the Orient*, *51*, 142–155. doi:10.2117/psysoc.2008.142
- Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*, 116–119. doi:10.1126/science.1111709
- Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar faces recognition: A review. *Memory*, *17*, 577–596. doi:10.1080/09658210902976969
- Kreiman, J., & Papcun, G. (1991). Comparing discrimination and recognition of unfamiliar voices. *Speech Communication*, *10*, 265–275. doi:10.1016/0167-6393(91)90016-M
- Leite, W. L., & Beretvas, S. N. (2005). Validation of scores on the Marlowe-Crowne Social Desirability Scale and the Balanced Inventory of Desirable Responding. *Educational and Psychological Measurement*, *65*, 140–154. doi:10.1177/0013164404267285

- Merckelbach, H., Jelicic, M., & Pieters, M. (2011). Misinformation increases symptom reporting—a test-retest experiment. *Journal of the Royal Society of Medicine*, *2*. doi:10.1258/shorts.2011.011062
- Mullennix, J. W., Ross, A., Smith, C., Kuykendall, K., Conard, J., & Barb, S. (2011). Typicality effects on memory for voice: Implications for earwitness testimony. *Applied Cognitive Psychology*, *25*, 29–34. doi:10.1002/acp.1635
- Olsson, N. (2000). A comparison of correlation, calibration, and diagnosticity as measures of the confidence-accuracy relationship in witness identification. *Journal of Applied Psychology*, *85*, 504–511. doi:10.1037//0021-9010.85.4.504
- Orchard, T. L., & Yarmey, A. D. (1995). The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology*, *9*, 249–260. doi:10.1002/acp.2350090306
- Sagana, A., Sauerland, M., & Merckelbach, H. (2012). *Choice blindness in eyewitness identifications for simultaneous lineups*. Manuscript in preparation.
- Sarno, J. A., & Alley, T. R. (1997). Attractiveness and the memorability of faces: Only a matter of distinctiveness? *American Journal of Psychology*, *110*, 81–92. doi:10.2307/1423702
- Schiller, N. O., & Köster, O. (1998). The ability of expert witnesses to identify voices: A comparison between trained and untrained listeners. *Forensic Linguistics*, *5*, 1–9.
- Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin*, *100*, 139–156. doi:10.1037/0033-2909.100.2.139
- Stevenage, S. V., Howland, A., & Tippelt, A. (2011). Interference in eyewitness and earwitness recognition. *Applied Cognitive Psychology*, *25*, 112–118. doi:10.1002/acp.1649
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology*, *43A*, 161–204.
- Wickham, L. H. V., & Morris, P. E. (2003). Attractiveness, distinctiveness, and recognition of faces: Attractive faces can be typical or distinctive but are not better recognized. *American Journal of Psychology*, *116*, 455–468. doi:10.2307/1423503
- Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and Aging*, *15*, 88–99. doi:10.1037/0882-7974.15.1.88

Received 21 November 2011; revised version received 10 January 2012