



Co-Editors: James W. Tanaka & Christian N.L. Olivers

Special issue:  
Person Perception  
Guest editors:  
Gadi Yovel, Galia Avidan, and Alice O'Toole



ISSN: 1350-6285 (Print) 1464-0716 (Online) Journal homepage: <http://www.tandfonline.com/loi/pvis20>


## Item consistency in retrieving person-specific semantic information from faces and voices: An exploratory study in healthy subjects

Chiara Piccininni, Guido Gainotti, Luigi Trojano, Simona Luzzi, Costanza Papagno, Giovanni Augusto Carlesimo, Camillo Marra & Davide Quaranta

To cite this article: Chiara Piccininni, Guido Gainotti, Luigi Trojano, Simona Luzzi, Costanza Papagno, Giovanni Augusto Carlesimo, Camillo Marra & Davide Quaranta (2017) Item consistency in retrieving person-specific semantic information from faces and voices: An exploratory study in healthy subjects, *Visual Cognition*, 25:4-6, 679-689, DOI: [10.1080/13506285.2017.1344340](https://doi.org/10.1080/13506285.2017.1344340)



To link to this article: <https://doi.org/10.1080/13506285.2017.1344340>

 View supplementary material 

 Published online: 31 Jul 2017.

 Submit your article to this journal 


 Article views: 28

 View related articles 

 View Crossmark data 



## Item consistency in retrieving person-specific semantic information from faces and voices: An exploratory study in healthy subjects

Chiara Piccininni<sup>a</sup>, Guido Gainotti<sup>a,b</sup>, Luigi Trojano <sup>c,d</sup>, Simona Luzzi<sup>e</sup>, Costanza Papagno<sup>f</sup>, Giovanni Augusto Carlesimo<sup>b</sup>, Camillo Marra<sup>a</sup> and Davide Quaranta<sup>a</sup>

<sup>a</sup>Research Center for Neuropsychology, Institute of Neurology, Catholic University of the Sacred Heart, Rome, Italy; <sup>b</sup>Laboratory of Clinical and Behavioural Neurology, IRCCS Santa Lucia Foundation, Rome, Italy; <sup>c</sup>Department of Psychology, Second University of Naples, Caserta, Italy; <sup>d</sup>Department of Neurorehabilitation, Maugeri Foundation, IRCCS, Telesse Terme, Italy; <sup>e</sup>Department of Clinical and Experimental Medicine, Polytechnic University of Marche, Ancona, Italy; <sup>f</sup>Department of Psychology, University of Milano-Bicocca, Milan, Italy

### ABSTRACT

Controversies exist over the format of person-specific semantic representations in healthy subjects and the loss of part of these representations in conditions of brain pathology. Some authors have suggested that in brain-damaged patients item-specific consistency of errors through different recognition modalities might indicate a loss of person-specific information. This view is hindered by the fact that item consistency might also be variable in healthy subjects, because names, faces and voices have different degrees of effectiveness in the recognition of familiar people. The aim of the present research was to try to clarify this issue by assessing the value of consistency of semantic retrieval in a large sample of Italian healthy subjects who had judged as familiar the faces and voices of 40 Italian celebrities. The effect of fame level on item consistency was also evaluated. Results showed that the degree of item consistency between faces and voices was rather variable, for example it was influenced by fame level when the latter was based on an integrated (non-verbal) face/voice familiarity score, but not when it was based on the celebrities' names in a (verbal) questionnaire. The consistency between non-verbal assessment of fame and item consistency in retrieving information from faces and voices suggests a different specialization of the left anterior temporal lobe (ATL) for the verbally coded aspects of personal semantics and of the right ATL for the visual (face) and auditory (voice) aspects of person representation.

### ARTICLE HISTORY



Received 8 June 2016  
Accepted 6 June 2017


### KEYWORDS

Personal semantics; face recognition; voice recognition; familiarity feelings; cross-modal item-specific consistency

In each social species the identification of individuals who belong to a social group and with whom we are personally acquainted or who are famous for their achievements is a fundamental biological function. For this reason, a complex multimodal recognition system has evolved in the brain to quickly and efficiently accomplish this difficult process of individual identification. Faces and voices are the most important stimuli in the visual and auditory domains through which we recognize familiar people and we interact with other persons, because they convey critical information not only about identity, but also about gender, age and emotional status of known and unknown people. Data recently gathered in the field of experimental social psychology have, however, shown that these two channels are not equally efficient in the identification of familiar people, because

faces play a more important role than voices in person recognition. Hanley, Smith, and Hadfield (1998) were the first authors to describe this asymmetry. They asked undergraduate students to identify a set of celebrities from either their voices or their faces. They found that the participants were better at recognizing celebrities from their faces than from their voices. Indeed 94% of the famous faces vs. 70% of the famous voices were considered as familiar; furthermore, among the stimuli judged as familiar, the recall of the corresponding occupation was correctly made by 92% of the participants in the face condition, but only by 63% in the voice condition. In other words, in the voice condition subjects gave significantly more “unfamiliar” responses and “familiar-only” responses than subjects in the face condition. These findings are consistent with the view that biographical

**CONTACT** Guido Gainotti  [guido.gainotti@unicatt.it](mailto:guido.gainotti@unicatt.it)  Research Center for Neuropsychology, Institute of Neurology, Catholic University of the Sacred Heart, Largo Agostino Gemelli, 8, 00168, Rome, Italy; Laboratory of Clinical and Behavioural Neurology, IRCCS Santa Lucia Foundation, Rome, Italy

 Supplemental data for this article can be accessed <https://doi.org/10.1080/13506285.2017.1344340>.

© 2017 Informa UK Limited, trading as Taylor & Francis Group

information is more easily associated with faces than with voices. Hanley and Turner (2000) tried to account for these unexpected findings by lowering recognition performance in the face condition to the same level as in the voice condition by presenting the faces out of focus. Under these circumstances, it proved just as difficult to recall the occupations of faces found familiar as it was to recall the occupations of voices found familiar. The authors argued that these results did not support the view that the voice pathway is substantially weaker than the face pathway. However, subsequent studies (Barsics & Brédart, 2011; Brédart, Barsics, & Hanley, 2009; Damjanovic & Hanley, 2007; Hanley & Damjanovic, 2009) confirmed and extended Hanley et al. (1998) results and showed that, even if the overall level of recognition is matched in the face and voice condition by presenting blurred faces, subjects still recall significantly less episodic and semantic information from familiar voices than from blurred familiar faces. Furthermore, Barsics and Brédart (2012) showed that the advantage of faces over voices can be confirmed by comparing participants' ability to associate semantic information with newly learned faces and voices in very well-controlled conditions.

The contrast between the results obtained by Hanley and Turner (2000) and by ensuing authors could be due to the fact that the vocal material created by Hanley and Turner (2000) included extracts in which celebrities were talking about topics leaking information about their identity, whereas the following authors tried to ensure that the speech content pronounced by the target person was devoid of any cues about their personal identity.

Thus, an advantage of faces over voices in terms of access to personal semantic and episodic information has been documented by all studies that used an appropriate methodology. This fact is of theoretical interest because it is at variance with an important cognitive model of familiar people recognition, i.e., the Interaction Activation and Competition (IAC) model, proposed by Burton, Bruce, and Johnston (1990) which draws on the first influential model of face recognition, i.e., that constructed by Bruce and Young (1986). Both models are based on the distinction between some lower-level perceptual processes, a locus of convergence of the output of these processes and a unitary store of higher level cognitive or semantic representations. The perceptual processes concern the visual and auditory channels through

which a seen face and a heard voice are mapped onto the corresponding invariant representation within specific recognition units for faces (FRUs) and voices (VRUs). The output of these modality-specific recognition units converges into person-identity nodes (PINs), which allow identification of a person characterized by a given face and voice, permitting access to the corresponding semantic (biographical) information. However, at variance with Bruce and Young's (1986) model, which assumes that PINs store semantic information, the IAC model maintains that PINs simply provide a modality-free gateway to a single semantic system, where information about people is stored in an amodal format. The results obtained by Hanley et al. (1998) and confirmed by Damjanovic and Hanley (2007), Brédart et al. (2009), Hanley and Damjanovic (2009) and Barsics and Brédart (2011, 2012) create problems for the IAC model. Specifically, if PINs provide a modality-free gateway to a single system that stores semantic information about people, and if familiar-only responses reflect a block between PINs and this semantic store, there would be no reason to expect that the number of familiar-only responses would be greater for voices than for degraded faces.

Data at variance with the IAC model have also been gathered in patients with brain damage and, in particular, in patients with right and left forms of Semantic Dementia (corresponding to the temporal variants of fronto-temporal degeneration, in which the atrophy prevails clearly in the right or in the left temporal lobes in the early stages of the disease). These data, which were reported by Snowden, Thompson, and Neary (2004), not only questioned the assumption that PINs provide a modality-free gateway to a person-specific semantic system, but also suggested that, instead of being stored in a unitary amodal format, personal semantic information is mainly based on visual (face) and auditory (voice) information in the right temporal lobe and on verbally-coded (name) information in the left temporal lobe (see Gainotti, 2007, 2015, for critical surveys).

Another method for checking the unitary amodal vs. the multimodal nature of material "stored" in the person-specific semantic system would be to assess the item concordance between retrieval of personal semantic information through faces and voices for stimuli considered as familiar in both modalities in patients with person recognition disorders.

Barton and Corrow (2016) rightly noted that, if the person-specific semantic system is unitary and amodal, then in patients with a loss of personal semantics we should find not only a deficit of similar severity in retrieving person-specific information from familiar faces and voices, but also an across modalities consistency for individual test items. Thus, a patient who is unable to identify a familiar person from his face should also be unable to do so from his/her voice. This criterion is interesting because it has been used to study general semantic disorders observed in aphasic patients (e.g., Butterworth, Howard, & McLoughlin, 1984; Jefferies & Lambon Ralph, 2006) and in patients with Semantic Dementia (e.g., Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000; Jefferies & Lambon Ralph, 2006) to distinguish disorders due to a loss of semantic information from those due to a retrieval or semantic control defect.

However, its practical application to the study of person-specific semantic disorders raises two kinds of problems. The first is to correct the consistency found in pathological patients with that usually observed in healthy subjects, because the different effectiveness of face and voice in the identification of famous people (documented in the first part of this introduction) suggests that item concordance might not be at ceiling in healthy subjects. The second problem concerns the possible influence of the fame level of each celebrity on the value of concordance between retrieval of his/her personal semantic information through face and voice stimuli. In the realm of personal semantics, fame level could, indeed, be considered as very similar, to the construct of “familiarity” in the domain of general semantics. Several authors (e.g., Bozeat et al., 2000; Funnell, 1995) have shown that in diseases characterized by a loss of semantic information, such as Semantic Dementia, the lost items are usually the less familiar ones, face-voice consistency might be higher in more famous than in less famous celebrities. Therefore, a study of the relationships between level of fame and degree of consistency between person-specific information retrieved on the same items from familiar faces and voices in healthy subjects could be of interest from two points of view. The first, more theoretical, study might be to evaluate whether in healthy subjects the across modalities consistency in retrieving person-specific information is or

is not influenced by the level of fame of the target people. The second, more applicative, study could be to provide researchers who are aiming to clarify the nature of person recognition disorders concerning faces and voices with normative data, which would allow them to correct the consistency values obtained in patients with multimodal people recognition disorders according to the corresponding values obtained by healthy subjects with people of the same fame level.

The two goals of the present investigation, which will be taken into account separately, are: (1) to assess in a sample of normal subjects item consistency in retrieving person-specific semantic information from faces and voices for a number of persons considered as famous or very famous in contemporary popular Italian culture; (2) to evaluate whether item consistency varies as a function of the fame level of these celebrities, in a separate group of normal subjects.

## **Assessment of item consistency in retrieving person-specific semantic information from faces and voices**

### ***Materials and methods***

To assess the item consistency in retrieving the person-specific semantic information about famous people from their faces and voices, we used the material of a new test battery: the Famous People Recognition Battery (FPRB). In this battery subjects are requested to recognize the same 40 persons (very well-known at the national level) through their faces and voices, distinguishing them from faces and voices of 20 unknown people (familiarity check), and to provide information indicating the correct identification of persons recognized as familiar. Identification from both faces and voices is assessed in the FPRB with a score ranging for each famous person between 0 (complete lack of knowledge) and 3 (very good knowledge). This identification score is based on three questions. The first two questions have a multiple-choice format and explore the general and specific occupational categories to which famous persons belong. The third question is open and subjects are asked to provide unequivocally identifying information about the person. One point is assigned to each correct answer. This section is administered

only for faces and voices that have been judged as familiar by the subject. The data necessary to directly compare responses obtained with face and voice stimuli have been recently reported by Quaranta et al. (2016) in a normative study conducted on 193 healthy subjects. All participants were community dwelling individuals, living independently. Thirty-eight of them were discarded on the basis of different criteria. Exclusion criteria were: educational level below three years of schooling; any current or prior neurological disease affecting CNS (e.g., brain injury or stroke); current or past history of alcohol or drug abuse; current depression or major psychiatric diseases; belonging to a family with a genetic form of dementia; chronic medical conditions potentially affecting CNS (e.g., hypothyroidism, renal or hepatic failure). Item consistency in retrieving person-specific semantic information from faces and voices was investigated in the remaining 155 healthy subjects.

Since results of the present investigation are based on data gathered in the normative study of the FPRB, we think necessary to summarize here the main features of this Battery.

#### ***The Famous People Recognition Battery***

Two main steps of the recognition process were considered in the construction of the FPRB. The first step consisted of considering as familiar or unfamiliar a given face and the corresponding voice. The second step consisted of providing information about people whose faces and voices were judged as familiar. Familiarity evaluation and person identification of 40 persons very well-known at the national level through their faces and voices were, therefore assessed by means of independent, parallel tests: the Familiarity evaluation and person identification from voices (VO-REC) and the Familiarity evaluation and person identification from faces (FA-REC).

***Familiarity evaluation and person identification from voices.*** The VO-REC was formed by 60 items (40 audio fragments of famous voices and 20 audio fragments of non-famous voices). All fragments were extracted from publicly available video or audio registrations and lasted about 15 s. Fragments of famous voices did not contain any information that could allow the direct recognition of the person (neutral discourses). The tracks were in .mp3 format, with compression at 320 kbps; the track intensity was

normalized; the peak intensity was  $-3$  db, with RMS of about  $-23$  db for each channel (the tracks were in stereo format); the sound pressure delivered through headphones was 60–65 dB. They were presented using high resolution headphones (Grado Prestige Series SR225e) through an audio player software (VLC Media Player) running under Windows 7 Professional Operating System; both the examiner and the subjects were hearing the track simultaneously. The test sessions were conducted in quiet room in which every possibly confounding element in the surroundings was removed (e.g., windows and curtains were shut). The stimuli were presented in a fixed order across subjects. Each subject was asked to carefully listen to the audio fragment and to provide a familiarity judgment. If the answer concerning the familiarity judgment was positive and the voice belonged to a famous person, subjects were asked three further questions. The first two questions had a multiple-choice format and explored the general and specific occupational categories to which famous persons belonged. The third question was open and the subject was asked to provide unequivocally identifying information about the person. One point was assigned to each correct answer (semantic score for each voice 0–3; total semantic score for voices 0–120).

***Familiarity evaluation and person identification from faces.*** The test was formed by black-and-white photos of the same famous persons who entered the VO-REC, along with 20 black-and-white photos of non-famous persons. The images were presented on a high resolution (1920 × 1200 pixel) monitor (Dell Ultrasharp U2415 LED monitor). They were preprocessed by means of Adobe Photoshop to obtain approximately the same contrast and brightness; the size was standardized, being 20 cm (vertical axis) × 14 cm (subtending approximately 28 × 20 degrees of visual angle at 40 cm viewing distance). Each image included just the face of the subject in frontal view, with a neutral expression; the background was artificially blurred during preprocessing. Each subject was asked to look at the photo and to provide a familiarity judgment. The procedure used to evaluate person identification from faces was the same used to evaluate person identification from voices.



### **Assessment of the possible role of guessing in FPRB**

The FPRB includes two multiple choice questions that could be possibly answered on the basis of perceptual cues (for example, a good-looking man could be identified as someone involved in entertainment). In order to evaluate the influence that facial appearance or vocal cues could have in predicting occupational categories, irrespectively of the specific knowledge about a famous person, we collected data on this subject in a subsample of 60 subjects, age-, gender- and education-matched to those on which data about semantic knowledge had been collected, asking them to guess the general and the specific occupational categories of famous persons that they had considered as unfamiliar. Few subjects were able to predict occupational categories for stimuli considered as unfamiliar; the mean occurrence of general and specific category guessing was 5.5% for VO-REC and 4.3% for FA-REC; detailed data for each stimulus are reported in Table S1 (Supplemental data).

### **Assessment of consistency**

The level of consistency between scores obtained in the face and voice modalities was assessed by determining the proportion of overlapping information that was retrieved in each modality for each famous person from voices and faces. Thus, the possible level of item consistency ranged from 0 (complete absence of consistency) to 1 (complete between modalities consistency). The third question of the semantic questionnaire asked the subjects for recalling as many information as possible that could lead to identification of the famous person; in this case consistency was considered to be reached if at least one of the features was recalled in both modalities. This approach was chosen since a great number of responses was possible for most of the famous persons evaluated, thus a complete overlap of information was predictably difficult to be reached. As an example, we can take the case of the actor Roberto Benigni. We consider a hypothetical subject who attributed the stimulus to the general occupational category "entertainment" and to the specific category "cinema" in both modalities (VO-REC and FA-REC); on the third question, on the FA-REC the same subject recalled the movie titles "La vita è bella", "Il piccolo diavolo" and "Non ci resta che piangere" on the VO-REC, stated that Benigni won the Academy Award for the best movie in 1999, and that

Benigni co-starred with Massimo Troisi in the movie entitled "Non ci resta che piangere". The concordance was quantified as follows: the general and specific occupational categories were the same, and at least one of the features recalled in the third question was the same; thus three out of three questions received the same response:  $3/3 = 1$ . We have chosen this method since it is quite conservative in nature, and an overestimation more than an underestimation of concordance was predictable.

The number of retrieved facts in common between the two modalities (i.e., facts identifier of a specific person retrieved in both the modalities) was also assessed by taking into account only the performance on the third question.

Only scores obtained by subjects who had experienced a familiarity feeling for both modalities of presentation were taken into account in the assessment of consistency.

### **Statistics**

Comparisons of frequencies were carried out by means of  $\chi^2$  with Yates' continuity correction and Fisher's exact test as requested. Mean comparisons were carried out using the *t*-test after Levene's test for equality of variances.

Since a large number of statistical analysis was carried out, the Bonferroni's correction was applied, and the significance level was set to 0.001.

### **Results**

#### **Comparison of the occurrence of familiarity feelings from face and voice for each of the 40 famous persons included in the FPRB**

The sample was composed of 87 women (56.1%) and 68 men, with mean age of 47.90 years ( $SD = 14.521$ ) and mean education of 12.33 years ( $SD = 4.597$ ). Table 1 displays the level of concordance in the production of a familiarity feeling from face and voice for each of the 40 famous persons included in the FPRB. In this preliminary assessment of the familiarity judgements for famous faces and voices we took into account the performance of all 155 healthy subjects. As shown, for most of the stimuli (26 out of 40) the familiarity feeling was more easily evoked from faces; for none of the stimuli we found an advantage for the presentation of voices.

**Table 1.** Comparison of the occurrence of the familiarity feeling between the two modalities of presentation of the stimuli (voices and faces) for each famous person included in the test battery. Number of subjects who reported a familiarity feeling for each stimulus when the presentation was made from faces or from voices, alongside with the corresponding percentage of the sample, is reported. Bold font indicates significant differences.

	Familiarity check from voices		Familiarity check from faces		$\chi^2$	<i>p</i>
	<i>N</i>	%	<i>N</i>	%		
Adriano Celentano	130	83.9	155	100	25.1	<b>&lt;.001</b>
Albano Carrisi	84	54.2	154	99.4	86.1	<b>&lt;.001</b>
Alberto Sordi	131	84.5	154	99.4	21.1	<b>&lt;.001</b>
Aldo Fabrizi	125	80.6	136	87.7	2.4	.119
Andrea Bocelli	82	52.9	101	65.2	4.3	.038
Antonella Clerici	95	61.3	150	96.8	56.8	<b>&lt;.001</b>
Benedetto XVI	133	85.8	147	94.8	6.2	.013
Beppe Grillo	119	76.8	155	100.0	38.5	<b>&lt;.001</b>
Bruno Vespa	152	98.1	155	100.0	1.3	.246
Corrado	144	92.9	152	98.1	3.7	.056
Enrico Mentana	118	76.1	155	100	39.8	<b>&lt;.001</b>
Fabrizio Frizzi	94	60.6	150	96.8	58.2	<b>&lt;.001</b>
Francesco Totti	119	76.8	151	97.4	27.6	<b>&lt;.001</b>
Gad Lerner	95	61.3	140	90.3	34.1	<b>&lt;.001</b>
Gianfranco Fini	84	54.2	151	97.4	76.6	<b>&lt;.001</b>
Giorgio Napolitano	132	85.2	153	98.7	17.4	<b>&lt;.001</b>
Giovanni Paolo II	113	72.9	155	100.0	46.3	<b>&lt;.001</b>
Lilli Gruber	101	65.2	148	95.5	43.2	<b>&lt;.001</b>
Luciana Littizzetto	152	98.1	154	99.4	0.3	.615
Mara Venier	146	94.2	153	98.7	3.4	.065
Mario Monti	96	61.9	154	99.4	67.1	<b>&lt;.001</b>
Massimo D'Alema	91	58.7	149	96.1	60	<b>&lt;.001</b>
Maurizio Costanzo	150	96.8	155	100.0	3.3	.071
Michele Santoro	98	63.2	143	92.3	36.1	<b>&lt;.001</b>
Mike Bongiorno	140	90.3	155	100	13.7	<b>&lt;.001</b>
Nichi Vendola	94	60.6	149	96.1	55.5	<b>&lt;.001</b>
Nino Manfredi	110	71.0	138	89.0	14.7	<b>&lt;.001</b>
Patty Pravo	94	60.6	139	89.7	33.5	<b>&lt;.001</b>
Pierferdinando Casini	138	89.0	151	97.4	7.4	.007
Pierluigi Bersani	95	61.3	154	99.4	68.7	<b>&lt;.001</b>
Piero Angela	134	86.5	142	91.6	1.6	.203
Pippo Baudo	125	80.6	154	99.4	28.1	<b>&lt;.001</b>
Raffaella Carrà	112	72.3	155	100.0	47.6	<b>&lt;.001</b>
Raimondo Vianello	149	96.1	155	100.0	4.2	.039
Renato Rascel	108	69.7	116	74.8	0.8	.375
Roberto Benigni	152	98.1	155	100.0	1.3	.246
Sandro Pertini	95	61.3	139	89.7	32.2	<b>&lt;.001</b>
Silvio Berlusconi	146	94.2	154	99.4	5.1	.024
Totò	122	78.7	154	99.4	31.7	<b>&lt;.001</b>
Umberto Bossi	123	79.4	153	98.7	27.8	<b>&lt;.001</b>

The mean number of false alarms (i.e., stimuli corresponding to non-famous people eliciting a familiarity feeling) was respectively 2.10 ( $SD = 2.898$ ) on VO-REC and 1.77 ( $SD = 2.703$ ) on FA-REC; there was no statistically significant difference between these values ( $|t|_{154} = 1.274, p = .205$ ).

**Comparison between the mean semantic scores obtained for each stimulus in the two modalities by subjects who had recognized both faces and voices as familiar**

In this study and in the assessment of consistency in retrieving the person-specific semantic information

about a celebrity, only data obtained by subjects who had recognized as familiar both his/her voice and his/her face were taken into account. The total semantic scores obtained when information was retrieved from voices (mean = 69.81;  $SD = 24.326$ ) and from faces (mean = 101.06;  $SD = 15.288$ ) were significantly different ( $|t|_{154} = 22.419; p < .001$ ). This difference is reflected in the semantic scores obtained by individual subjects (see Table 2 where the comparison between the mean semantic scores obtained for each stimulus in the two modalities is reported).

In most cases the semantic scores obtained when the stimuli were presented via the auditory modality (voices) were significantly lower than the

**Table 2.** Comparison of the semantic score obtained for each stimulus when the presentation was performed from voice and face respectively. Bold font indicates significant differences. *SD* = standard deviation.

	Semantic score from voices		Semantic score from faces		$ t $	<i>p</i>
	Mean	<i>SD</i>	Mean	<i>SD</i>		
Adriano Celentano	2.39	0.976	2.88	0.341	5.47	<b>&lt;.001</b>
Albano Carrisi	2.33	1.079	2.93	0.283	4.96	<b>&lt;.001</b>
Alberto Sordi	2.53	0.727	2.84	0.459	4.32	<b>&lt;.001</b>
Aldo Fabrizi	2.23	0.908	2.52	0.666	2.92	.004
Andrea Bocelli	1.46	1.229	2.76	0.619	8.71	<b>&lt;.001</b>
Antonella Clerici	2.31	0.979	2.85	0.38	5.15	<b>&lt;.001</b>
Benedetto XVI	2.55	0.763	2.65	0.775	1.059	.291
Beppe Grillo	2.09	1.150	2.5	0.893	3.17	.002
Bruno Vespa	2.72	0.767	2.87	0.336	2.27	.024
Corrado	2.31	1.067	2.84	0.467	5.41	<b>&lt;.001</b>
Enrico Mentana	2.33	0.952	2.61	0.551	2.88	.005
Fabrizio Frizzi	2.64	0.716	2.81	0.469	2.1	.037
Francesco Totti	2.48	1.040	2.85	0.538	3.513	<b>&lt;.001</b>
Gad Lerner	1.97	1.224	2.36	0.702	2.85	.005
Gianfranco Fini	2.33	0.883	2.32	0.844	0.131	.896
Giorgio Napolitano	2.08	1.089	2.52	0.679	4.082	<b>&lt;.001</b>
Giovanni Paolo II	2.44	0.876	2.77	0.439	3.63	<b>&lt;.001</b>
Lilli Gruber	2.10	1.005	2.5	0.742	3.42	<b>&lt;.001</b>
Luciana Littizzetto	2.76	0.498	2.73	0.499	0.52	.606
Mara Venier	2.61	0.699	2.79	0.468	2.62	.009
Mario Monti	1.85	1.105	2.24	0.784	2.987	.003
Massimo D'Alema	1.52	1.320	2.30	0.786	5.148	<b>&lt;.001</b>
Maurizio Costanzo	2.69	0.795	2.91	0.288	3.24	<b>&lt;.001</b>
Michele Santoro	2.04	1.083	2.47	0.637	3.51	<b>&lt;.001</b>
Mike Bongiorno	2.53	0.869	2.97	0.177	5.87	<b>&lt;.001</b>
Nichi Vendola	2.12	1.226	2.40	0.876	1.919	.057
Nino Manfredi	1.77	0.964	2.54	0.663	7.15	<b>&lt;.001</b>
Patty Pravo	1.19	1.346	2.43	0.852	7.92	<b>&lt;.001</b>
Pierferdinando Casini	1.89	1.016	2.18	0.865	2.579	.010
Pierluigi Bersani	1.88	1.254	2.32	0.799	3.062	.0030
Piero Angela	2.76	0.578	2.69	0.746	0.89	.375
Pippo Baudo	2.51	0.972	2.86	0.344	3.85	<b>&lt;.001</b>
Raffaella Carrà	2.26	0.947	2.79	0.466	5.51	<b>&lt;.001</b>
Raimondo Vianello	2.61	0.723	2.85	0.413	3.45	<b>&lt;.001</b>
Renato Rascel	2.05	0.970	2.15	0.816	0.83	.405
Roberto Benigni	2.88	0.387	2.91	0.33	0.84	.399
Sandro Pertini	1.31	1.158	2.17	0.96	5.971	<b>&lt;.001</b>
Silvio Berlusconi	2.84	0.494	2.88	0.322	0.839	.402
Totò	2.34	0.986	2.84	0.381	5.3	<b>&lt;.001</b>
Umberto Bossi	1.81	1.074	2.50	0.736	6.074	<b>&lt;.001</b>

corresponding ones obtained in the visual modality (faces). Even in this case, none of the stimuli showed an advantage of voices.

### *Levels of concordance between the semantic scores obtained in the face and voice modality*

The mean levels of concordance for each stimulus are reported in Table 3. As predicted, the concordance level was above chance (one-sample *t*-test with comparison to zero:  $|t|_{39}=43.663$ ;  $p < .001$ ). As shown, the level of concordance is quite variable, ranging from very high values (e.g., 0.942 in the case of the actor Roberto Benigni and 0.931 in the case of the politician Silvio Berlusconi) to low values (e.g., 0.461 in the case of the singer and show girl Patty Pravo), with a

**Table 3.** Mean concordance between the semantic scores and mean number of facts retrieved from both the modalities of presentation. *N* = number of subjects who retrieved a familiarity feelings in both the auditory (voices) and visual (faces) modalities.

	<i>N</i>	Mean concordance	Number of common facts
Adriano Celentano	130	0.808	1.14
Albano Carrisi	83	0.777	1.46
Alberto Sordi	130	0.871	1.71
Aldo Fabrizi	116	0.807	1.34
Andrea Bocelli	59	0.589	2.00
Antonella Clerici	93	0.769	1.59
Benedetto XVI	128	0.849	1.94
Beppe Grillo	119	0.694	1.41
Bruno Vespa	152	0.890	1.64
Corrado	141	0.785	1.28
Enrico Mentana	118	0.819	1.86
Fabrizio Frizzi	93	0.873	2.35
Francesco Totti	117	0.823	2.07
Gad Lerner	89	0.707	1.54
Gianfranco Fini	82	0.846	1.15
Giorgio Napolitano	130	0.823	0.98
Giovanni Paolo II	113	0.831	3.26
Lilli Gruber	96	0.707	1.24
Luciana Littizzetto	151	0.934	1.37
Mara Venier	144	0.888	1.37
Mario Monti	96	0.709	1.36
Massimo D'Alema	87	0.552	1.11
Maurizio Costanzo	150	0.885	2.19
Michele Santoro	95	0.777	1.44
Mike Bongiorno	140	0.845	1.97
Nichi Vendola	92	0.725	1.28
Nino Manfredi	100	0.655	0.52
Patty Pravo	82	0.461	0.74
Pierferdinando Casini	135	0.722	0.51
Pierluigi Bersani	94	0.669	0.20
Piero Angela	128	0.892	2.15
Pippo Baudo	125	0.831	1.46
Raffaella Carrà	112	0.763	0.79
Raimondo Vianello	149	0.871	1.69
Renato Rascel	91	0.780	2.03
Roberto Benigni	152	0.942	2.43
Sandro Pertini	88	0.528	1.92
Silvio Berlusconi	145	0.931	1.69
Totò	122	0.772	1.64
Umberto Bossi	122	0.671	1.48

median value of 0.796. The mean number of facts in common between the modalities was also above chance level (one-sample *t*-test with comparison to zero:  $|t|_{39}=16.834$ ;  $p < .001$ ) and varied broadly across stimuli, ranging from very low values (e.g., 0.2 in the case of politician Pierluigi Bersani) to high values (e.g., 3.26 in the case of former pope Giovanni Paolo II).

### **Assessment of the influence that the “fame level” can have on the item concordance**

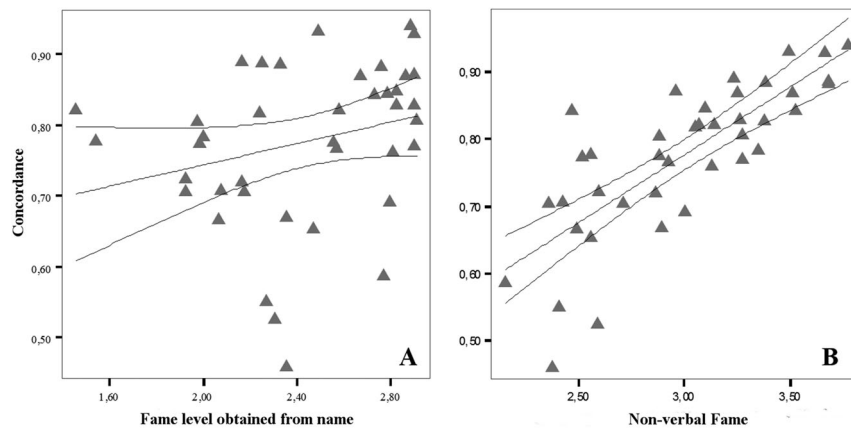
The degree of fame of the 40 famous persons whose faces and voices had been included in the FPRB was assessed by administering their names to a sample of 79 healthy subjects, age-, gender- and education-matched to those on which data about semantic knowledge were collected. For each person included in the FPRB, a score ranging from 0 (completely unknown) to 3 (very famous) was obtained. The mean score obtained for each person was used as an estimator of the level of fame. Furthermore, since this method of scoring fame was a verbal one, whereas the item concordance had been assessed between two non-verbal person identification modalities (face and voice), we also evaluated fame with a non-verbal fame score of the familiarity levels obtained from face and voice of each famous person by the subjects who participated in the first experiment. This was possible because, even though we had not formally rated familiarity for faces and voices with a score ranging from 0–3, we had noted all instances in which responses about familiarity for faces or voices had been given after a latency or with perplexity. Familiarity was rated 1 in these cases and 2 if responses had been rapid and sure. Summing up scores obtained through face and voice, we obtained for each famous person a score ranging from 2 (perplexity from both face and voice = low familiarity) to 4 (high familiarity), which reflected their implicit non-verbal evaluation of the person fame.

## **Results**

### *Influence that the verbal method of assessing fame could have on the item concordance between face and voice*

The mean level of fame attributed to each stimulus ranged from 1.46 to 2.91, with a mean value of 2.44 ( $SD = 0.395$ ).





**Figure 1.** Regression models in which the level of concordance was set as dependent variable, and (A) verbal and (B) non-verbal fame were set as predictors.

When entered into a linear regression model (Figure 1A) this score did not significantly predict the level of concordance between the two person identification modalities ( $F_{1,38} = 2.857$ ;  $p = .099$ ), accounting for less than 5% of the total sample variance ( $\text{Adj}R^2 = 0.045$ ); even in the case of facts retrieved in both modalities, the level of fame as assessed by a verbal method was not predictive ( $F_{1,38} = 1.840$ ;  $p = .183$ ;  $\text{Adj}R^2 = 0.020$ ) (Figure 2).

#### *Influence that the non-verbal, integrated familiarity method of assessing fame could have on the item concordance between face and voice*

The mean level of non-verbal fame ranged from 2.15 to 3.47, with a mean value of 3.00 ( $SD = 0.435$ ). The regression analysis (Figure 1B) showed that it significantly predicted the level of concordance between the modalities ( $F_{1,38} = 59.953$ ;  $p < .001$ ), accounting for about 60% of the total variance. The level of non-verbal fame also significantly predicted the number of facts retrieved in both modalities ( $F_{1,38} = 7.116$ ;  $p = .011$ ), accounting for about 14% of total variance ( $\text{Adj}R^2 = 0.136$ ).

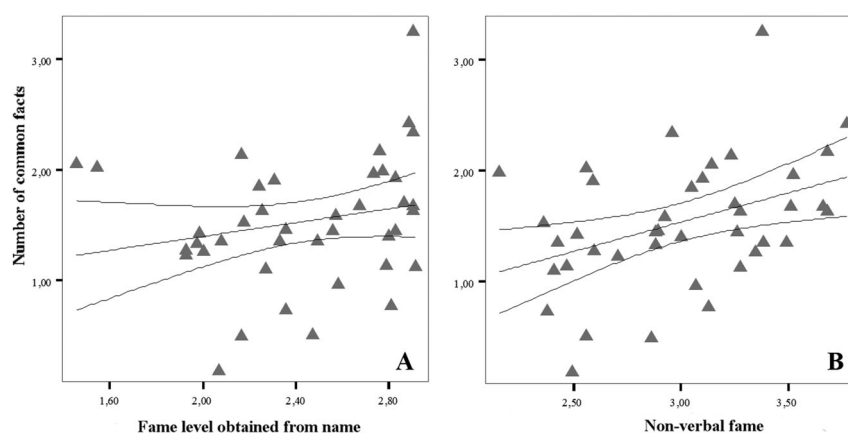
The effect of the non-verbal fame was also assessed by comparing the level of concordance across the groups of subjects with different familiarity level. The weighted mean and variance of concordance level, as well as the number of facts retrieved in both modalities, were obtained for subjects who obtained, respectively, a familiarity level of 2, 3 or 4. The weighted mean of concordance for subjects with familiarity 2, 3 or 4 were respectively 0.446 ( $SD = 0.0464$ ), 0.510 ( $SD = 0.0345$ ) and 0.940 ( $SD = 0.0252$ ). An

ANOVA with the level of non-verbal familiarity set as categorical predictor yielded a significant general effect ( $F_{1,38} = 2228.12$ ;  $p < .001$ ); post-hoc pairwise comparisons showed that the three level of familiarity differed significantly from each other ( $p < .001$  for all comparisons). The weighted mean of number of facts retrieved in common in the two modalities by subjects with familiarity 2, 3 or 4 were 0.02 ( $SD = 0.010$ ), 0.95 ( $SD = 0.009$ ) and 2.05 ( $SD = 0.129$ ); the ANOVA showed a significant general effect of non-verbal familiarity ( $p < .001$ ); all the pairwise post-hoc comparisons showed statistically significant differences ( $p < .001$  for all comparisons).

#### **General discussion**

The main results of the present study can be summarized as follows:

- With respect to the asymmetry between face and voice in the identification of famous people, our results, conducted on 155 healthy subjects, using a newly developed Famous People Recognition Battery (FPRB) tailored for Italian participants, replicated existing work, showing that healthy subjects experience greater familiarity from faces than voices, and obtain a superior retrieval of semantic information from face than from voice for targets where both modalities are recognized as familiar.
- As for the level of concordance between the semantic scores obtained in the face and voice modalities, our results showed that they are quite variable, ranging from very high to very



**Figure 2.** Regression models in which the number of facts retrieved from both voice and face presentation was set as dependent variable, and (A) verbal and (B) non-verbal fame were set as predictors.

low scores, and this variability of item consistency seems in part related to the fame of the famous person.

- c) In order to clarify if the influence of fame level on item consistency could be related to the concordance between the (verbal or non-verbal) procedure used to evaluate fame and the non-verbal nature of the person identification modalities through which item consistency had been assessed, we used a verbal and a non-verbal method of fame evaluation. The verbal assessment of fame did not significantly predict the item consistency between the semantic scores obtained in the face and voice modality, whereas the non-verbal overall assessment of familiarity accounted for a large part of this item consistency.

Each of these main results will be more thoroughly discussed in the following paragraphs.

- a) Results obtained in our preliminary assessment of familiarity judgements for famous faces and voices and in the evaluation of person-specific semantic information obtained from faces and voices considered as familiar are consistent with previous literature data (Barsics & Brédart, 2011, 2012; Brédart et al., 2009; Damjanovic & Hanley, 2007; Hanley & Damjanovic, 2009; Hanley et al., 1998) showing that famous faces are more frequently recognized as familiar than the corresponding voices and that the amount of semantic information retrieved from faces is greater than the corresponding amount retrieved

from voices. These findings are not consistent with Burton et al.'s (1990) account of PIN as the point at which a person is recognized as familiar and the connections are established with the unitary semantic system, where information about people is stored in an amodal format. However, they are not necessarily at odds with an IAC model in which activation at PIN acts as a graded signaller of familiarity. Since the face is the dominant cue to identity (perhaps because celebrities are more often encountered in the media via their faces than their voices), it might activate PIN more than the voice, and thus might more successfully activate the semantic information accessed via the modality-free gateway (whether in a unimodal store or in multimodal stores in each hemisphere). In this regard, the data using blurred faces (vs. voices) are at odds with the IAC model, but the present data are not. It must be acknowledged that literature using personally familiar or newly learned stimuli in which relative exposure to face and voice are well controlled (e.g., Barsics, 2014; Stevenage, Hugill, & Lewis, 2012) suggest that the relative exposure argument is not necessarily the best way to account for the better recognition of faces than voices and that the face might be the more dominant cue for some other reason (i.e., expertise when discriminating between one face and another).

- b) The high variability obtained in the study of the consistency between the semantic scores obtained through face and voice suggests that in the evaluation of patients with impaired

knowledge of famous persons the study of item consistency must be cautiously considered. Part of this variability might, indeed, be related to the level of fame/familiarity of the person presented. This claim is consistent with results obtained by Damjanovic and Hanley (2007), who instructed their healthy subjects to rate the familiarity of a face or voice on a 0–3 scale and requested the person's occupation and name following a familiarity response of 1, 2 or 3. There were significantly more false alarms when the familiarity level was low and these familiarity effects influenced the retrieval of person-specific semantic information, because more occupations were correctly recalled when the face or voice had been judged of high familiarity, than when it had been found to be of medium or low familiarity.

- c) Two accounts can be given of the discrepancy observed between results obtained using the verbal questionnaire and the non-verbal fame assessment to evaluate the influence of fame on item consistency. The first is that retrieval of person-specific semantic information relies on the personal knowledge/familiarity that each person has for famous individuals, rather than on their shared fame. The second is that a concordance should exist between the modalities through which fame is assessed and the modalities of person recognition considered in the evaluation of item consistency. In this case the level of fame should influence the level of consistency when both are assessed through the same perceptual modalities, but not when fame is assessed through a verbal questionnaire and consistency concerns data acquired through non-verbal (visual and auditory) modalities.

In the first part of this paper we have considered data obtained in Semantic Dementia patients by Snowden et al. (2004) and by Gainotti (2007), which have suggested that personal semantic information, instead of being stored in a unitary, amodal format in both anterior temporal lobes (ATLs), could be mainly based on visual (face) and auditory (voice) information in the right ATL and on verbally-coded (name) information in the left ATL. This model, assuming a different specialization of the left ATL for the verbally coded aspects of personal semantics and of the right ATL for the visual (face) and auditory

(voice) aspects of person representation, has been confirmed by data obtained by von Kriegstein, Kleinschmidt, Sterzer, and Giraud (2005) and by von Kriegstein and Giraud (2006), in experiments which have shown that a cross-communication between face and voice channels of person recognition probably exists before the level of PINs. These authors measured by means of fMRI brain activity during voice identification tasks, in which subjects focused on either the speaker's voice or the verbal content of sentences, and showed that familiar persons' voices activated the Fusiform Face Area when the identification task was to focus on the speaker's identity. Schweinberger, Herholz, and Stief (1997) and O'Mahony and Newell (2012) have shown, however, that an interaction similar to that found between faces and voices is not observed between faces and names. These results suggest that the link between face and voice is closer than that between faces/voices on one hand and names on the other hand. The reason could be that the right hemisphere channels which process perceptual data are more closely integrated than the right and left hemisphere structures processing perceptual and verbal data, respectively. The same model could help to understand why a verbally assessed judgment of fame accounts for a very low portion of variance observed on item consistency in the retrieval of person-specific semantic information from face and voice, because this fame assessment, being strictly verbal in nature, was probably mainly subsumed by the left ATL. If at the end of this discussion we come back to the problem of how consistency of semantic retrieval across modalities in healthy subjects can help to understand the nature of semantic deficit retrieval in clinical populations, we can say that the study of item consistency is a feasible, but difficult task. This statement stems from different reasons: (a) the voice pathway is substantially weaker than the face pathway; (b) consistency across modalities is influenced by fame level, which is highly variable even within a sample of celebrities and is different from country to country and from generation to generation; (c) this influence is significant only when fame is assessed in terms of personal familiarity through the same visual and auditory modalities which are used for the assessment of item consistency, but not when fame is assessed by means of a verbal questionnaire. So, if we consider this issue in terms of costs/benefits, we don't think that the study

of item consistency may be considered as a benchmark to clarify the nature of semantic deficit retrieval in clinical populations.

## Acknowledgements

The authors would like to thank Dr Antonia Ferrara, Dr Natascia Guarino and Dr Giovanna Masone-Jacobucci for their precious help in data gathering.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## ORCID

Luigi Trojano  <http://orcid.org/0000-0002-0328-9642>

## References

- Barsics, C. (2014). Person recognition is easier from faces than from voices. *Psychologica Belgica*, *54*, 244–254.
- Barsics, C., & Brédart, S. (2011). Recalling episodic information about personally known faces and voices. *Consciousness and Cognition*, *20*, 303–308.
- Barsics, C., & Brédart, S. (2012). Recalling semantic information about newly learned faces and voices. *Memory*, *20*, 527–534.
- Barton, J. S., & Corrow, S. L. (2016). Recognizing and identifying people: A neuropsychological review. *Cortex*, *75*, 132–150.
- Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*, *38*, 1207–1215.
- Brédart, S., Barsics, C., & Hanley, J. R. (2009). Recalling semantic information about personally known faces and voices. *European Journal of Cognitive Psychology*, *21*, 1013–1021.
- Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327.
- Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, *81*, 361–380.
- Butterworth, B., Howard, D., & McLoughlin, P. (1984). The semantic deficit in aphasia: The relationship between semantic errors in auditory comprehension and picture naming. *Neuropsychologia*, *22*, 409–426.
- Damjanovic, L., & Hanley, J. R. (2007). Recalling episodic and semantic information about famous faces and voices. *Memory and Cognition*, *35*, 1205–1210.
- Funnell, E. (1995). Objects and properties: A study of the breakdown of semantic memory. *Memory*, *3*, 497–518.
- Gainotti, G. (2007). Different patterns of famous people recognition disorders in patients with right and left anterior temporal lesions: A systematic review. *Neuropsychologia*, *45*, 1591–1607.
- Gainotti, G. (2015). Implications of recent findings for current cognitive models of familiar people recognition. *Neuropsychologia*, *77*, 279–287.
- 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.
- 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.
- Hanley, J. R., & Turner, J. M. (2000). Why are familiar-only experiences more frequent for voices than for faces? *Quarterly Journal of Experimental Psychology*, *53A*, 1105–1116.
- Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. *Brain*, *129*, 2132–2147.
- O'Mahony, C., & Newell, F. N. (2012). Integration of faces and voices, but not faces and names, in person recognition. *British Journal of Psychology*, *103*, 73–82.
- Quaranta, D., Piccininni, C., Carlesimo, G. A., Luzzi, S., Marra, C., Papagno, C., ... Gainotti, G. (2016). Recognition disorders for famous faces and voices: A review of the literature and normative data of a new test battery. *Neurological Sciences*, *37*, 345–352.
- Schweinberger, S. R., Herholz, A., & Stief, V. (1997). Auditory long term memory: Repetition priming of voice recognition. *The Quarterly Journal of Experimental Psychology, Section A-Human*, *50*, 498–517.
- Snowden, J. S., Thompson, J. C., & Neary, D. (2004). Knowledge of famous faces and names in semantic dementia. *Brain*, *127*, 860–872.
- Stevenage, S. V., Hugill, A., & Lewis, H. G. (2012). Integrating voice recognition into models of person perception. *Journal of Cognitive Psychology*, *24*, 409–419.
- von Kriegstein, K., & Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology*, *4* (10), e326. doi:10.1371/journal.pbio.0040326
- von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience*, *17*, 367–376.