Spontaneous voice–face identity matching by rhesus monkeys for familiar conspecifics and humans

Julia Sliwa[a], Jean-René Duhamel[a], Olivier Pascalis[b], and Sylvia Wirth[a,1]

[a]Centre de Neuroscience Cognitive, Centre National de la Recherche Scientifique, Université Lyon I, 69675 Bron, France; and [b]Laboratoire de Psychologie et Neurocognition, Centre National de la Recherche Scientifique, Université Pierre Mendès France, 38040 Grenoble, France

Edited by Nikos K. Logothetis, Max Planck Institute for Biological Cybernetics, Tuebingen, Germany, and approved December 15, 2010 (received for review June 15, 2010)

Recognition of a particular individual occurs when we reactivate links between current perceptual inputs and the previously formed representation of that person. This recognition can be achieved by identifying, separately or simultaneously, distinct elements such as the face, silhouette, or voice as belonging to one individual. In humans, those different cues are linked into one complex conceptual representation of individual identity. Here we tested whether rhesus macaques (*Macaca mulatta*) also have a cognitive representation of identity by evaluating whether they exhibit cross-modal individual recognition. Further, we assessed individual recognition of familiar conspecifics and familiar humans. In a free preferential looking time paradigm, we found that, for both species, monkeys spontaneously matched the faces of known individuals to their voices. This finding demonstrates that rhesus macaques possess a cross-modal cognitive representation of individuals that extends from conspecifics to humans, revealing the adaptive potential of identity recognition for individuals of socioecological relevance.

cross-species | vocal communication | nonhuman primates | picture recognition

In humans, both faces and voices convey information about identity, providing some of the many cues we use to recognize individuals we know (1). The multifaceted nature of identity code suggests that a complex cognitive representation binds semantic information with information of different sensory modalities. In rhesus monkeys, however, it has typically been assessed only via single-modality information. For instance, rhesus monkeys can discriminate between calls of two conspecifics in a playback experiment using a spontaneous habituation–discrimination paradigm (2). They are also able to discriminate faces of two conspecifics in a match-to-sample task (3) and monkey faces or human faces in visual paired comparison tasks (4–6). These observations demonstrate that monkeys can *discriminate* idiosyncratic characteristics ("individual A is different from B") for their own or other species. However, they do not provide evidence of *individual recognition* ("this is individual A, this is B"). In comparison with discrimination, individual recognition requires an additional associative level that allows retrieval of information belonging to a specific individual. In rhesus monkeys, coarse recognition processes such as of their own species, kin, gender, reproductive status, or hierarchy are well documented (2, 7–12). However, these rudimentary recognition abilities fail to account for some sophisticated behaviors that are observed in rhesus macaques' societies. In particular, rhesus macaques live in large groups and maintain elaborate social relations involving, e.g., nonkin alliances during aggressive interactions, fight interference, reciprocal support, friendly grooming, and reconciliation (13–15). Such an organization would benefit well from the finest-grain individual recognition based on a cognitive multimodal representation of identity.

Individual recognition can be revealed in animals by demonstrating the existence of a cognitive representation of individuals supported by the integration of cues from multiple sensory modalities. Only a few studies have applied this approach to investigate individual recognition in species other than humans and chimpanzees (16, 17) and none of them focused on rhesus monkeys. Moreover, these studies explored individual recognition of peers (18, 19), but did not examine whether this capacity extended to other species. Given the complexity of individual recognition, the question of whether it can be applied to both one's own and another species remains open. When individual recognition was tested across species, e.g., dogs recognizing humans (20) or squirrel monkeys recognizing humans (21), the studies remained inconclusive regarding individual voice identification and failed to demonstrate face–voice association for more than one highly familiar human individual. Thus, the species-specific nature of individual recognition is still unclear. Yet, the cross-species issue is of special interest because it provides information about the properties of individual recognition by testing its adaptability. For instance, in human infants, discrimination of faces is known to specialize during the first year of life for conspecifics or even more drastically for individuals of their own socio-ethnic group (22–25). However, this faculty can be maintained with longer exposure to the other species or groups (e.g., ref. 26). Similarly in rhesus monkeys, performance at processing faces is poorer for other species than for their own (4, 6), but this asymmetry can be reversed in favor of humans if infant monkeys first experience human faces (27). Thus, it is of great interest to evaluate whether the multimodal process of individual recognition also generalizes to highly relevant information acquired throughout life, such as information about human individuals for laboratory monkeys.

In the present study, we aimed to evaluate both the individual recognition of conspecifics and the individual recognition of humans by adult rhesus monkeys. Our subjects had daily exposure to both rhesus monkey and human individuals from infancy and were familiarized with both the humans and other rhesus monkeys serving as stimuli in our experiment via recent real life daily exposure (housing "roommates," caregivers, and researchers). Individual recognition was then investigated through a cross-modal approach. Our experiment assessed whether rhesus macaques would spontaneously match two attributes of familiar individuals, i.e., match a voice to the photograph of its corresponding face. In this design, voice–face matching is supported by memories of previous interactions. This process entails two components. First, can monkeys match a voice with an appropriate pictorial content? And concomitantly, is the picture of a face sufficient to elicit recall of the appropriate individual? These observations will substantially contribute to our knowledge of rhesus macaques' aptitude for multimodal representations of others. Only affiliation calls such as coos and grunts were used as representing monkey-sound stimuli because their formant structure is a reliable marker

Author contributions: J.S., J.-R.D., O.P., and S.W. designed research; J.S. and S.W. performed research; J.S. analyzed data; and J.S., J.-R.D., O.P., and S.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: swirth@isc.cnrs.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1008169108/-/DCSupplemental.

of individual identity (28). Human speech was used as human-sound stimuli (Fig. 1*A*). Still color photographs of known individuals (Fig. 1*A*) were used. We did not use movies, because indexical matching on the basis of the redundancy between vocal-tract resonance and its physical shape is possible when face stimuli are presented for real (physical presence) or on a forward-played movie, whereas it is impossible when presented as a still image or on a reversed-played movie (29, 30).

Six subjects were tested in a free scanning task on the basis of a preferential looking paradigm. After hearing a voice sample while fixating on a central spot of light, the monkey freely explored a pair of face pictures, only one of which matched the voice (Fig. 1*B*). Eye movements were restrained to a virtual window around the pair of images represented by the black area in Fig. 1*B* and the subject was rewarded for maintaining its gaze within this exploratory window during the stimulus presentation. Importantly, the animal was not reinforced to gaze at one picture over the other. We then compared the preferential-looking time. We predicted that, if monkeys have a cross-modal representation of known individuals, they would preferentially attend to the picture matching the voice.

Results

Do Monkeys Spontaneously Look Longer at a Face When It Matches a Voice Previously Heard? To assess whether animals linked the heard voice sample to one of the face photographs, we examined looking times for all face pictures depending on their congruence with the preceding voice (Fig. 2*A*). This congruence accounts for the intrinsic variability in looking times across faces that may bias the time allocated to each of the two images. Looking behavior can be influenced by factors such as dominance or sexual coloration of the pictured individuals (9, 12). We confirmed that our pictures yielded different looking times with a one-way analysis of variance (ANOVA) with face identity as a factor (Table S1 and Fig. 3). By comparing looking time for a face preceded by its matching voice to that for the same face when it was preceded by a different voice, we captured the effect of voice–face concordance on the time allocated to each image. Results for each face and each pair were concatenated and the group difference was then compared with zero (absence of voice–face concordance effect). This average difference was significantly greater than zero [$t(103) = 4.48$, $P = 9 \times 10^{-6}$], revealing that animals look significantly longer at a face when it is preceded by its voice than when it is preceded by a different voice (Fig. 2*B*). This finding was replicated when we tested whether the proportion of time spent looking at the matching face, relative to the other face in the pair, was greater than expected by chance, despite differences in looking times for individual pictures (*SI Materials and Methods*, *Text S1*). Furthermore, we ensured that this effect was not driven by gender matching rather than identity matching (*SI Materials and Methods*, *Text S2*).

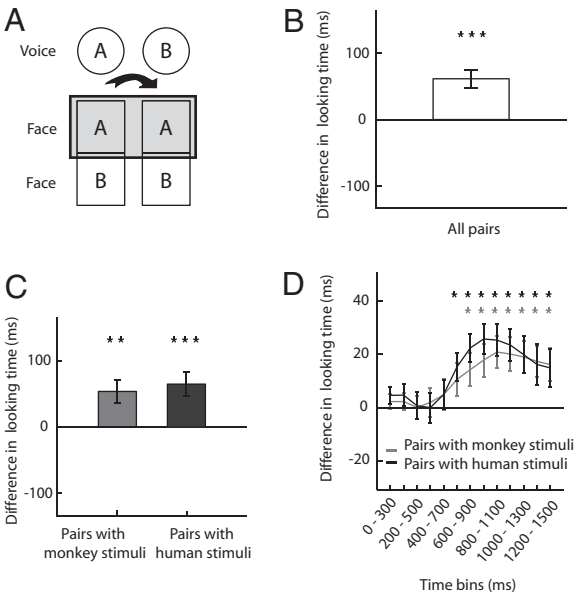

Fig. 2. Monkeys match the identity of the voice with the identity of the face. (*A*) For each pair AB, the box represents the relevant comparison: difference in looking time for face A when preceded by the congruent voice A compared with when preceded by the noncongruent voice B. (*B*) Mean looking time difference between the congruent condition and the noncongruent condition for all of the pairs ($n = 104$ pairs). The abscissa indicates chance expectation. Errors bars represent SEM (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). (*C*) The same as *B* but separately analyzed for stimulus pairs with monkey ($n = 32$ pairs) and human ($n = 72$ pairs). (*D*) The same as *C* but analyzed across time with a sliding window of 300 ms moved by100-ms steps for stimulus pairs with monkey ($n = 32$ pairs for each time point) and human ($n = 72$ pairs for each time point) (*$P < 0.05$; see Table S2 for exact *P* values).

Do Monkeys Exhibit a Cross-Modal Recognition of Both Conspecifics and Humans? The looking time for a given face when it was preceded by its matching voice, compared with when it was not, was significantly different for both monkey stimuli trials [$t_M(31) = 3.18$, $P = 0.0017$] and human stimuli trials [$t_H(71) = 3.52$, $P = 0.0004$] (Fig. 3*C*).

Change Over Time. To evaluate the point in time relative to the picture onset, at which animals started to look significantly more at the face matching the voice, we analyzed looking times obtained in sliding 300-ms windows moving in 100-ms steps. A significant effect was observed only from 800 ms after picture onset until the end of the visual presentation for trials with human stimuli and 900 ms for trials with monkey stimuli (Table S2 and Fig. 3*D*). Surprisingly, the voice–face concordance appears as a late effect.

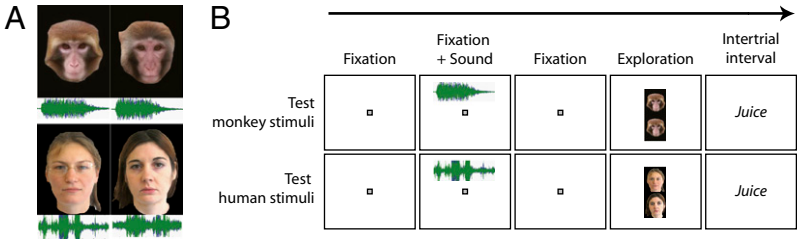

Fig. 1. Stimuli and experimental paradigm. (*A*) Examples of test stimuli for monkey (*Upper*) and human (*Lower*). Green diagrams represent spectrograms of "coo" vocalization and of voice samples of the individuals. (*B*) Test trials: The voice of a known individual is followed by the presentation of two pictures of known faces, one of which matches the preceding voice. The animals fixate for the 2 s during which the voice sample is played and explore freely during the picture presentation (1.5 s). Animals were rewarded for maintaining gaze within the boundary of the virtual window represented by the black area, regardless of their exploration pattern.

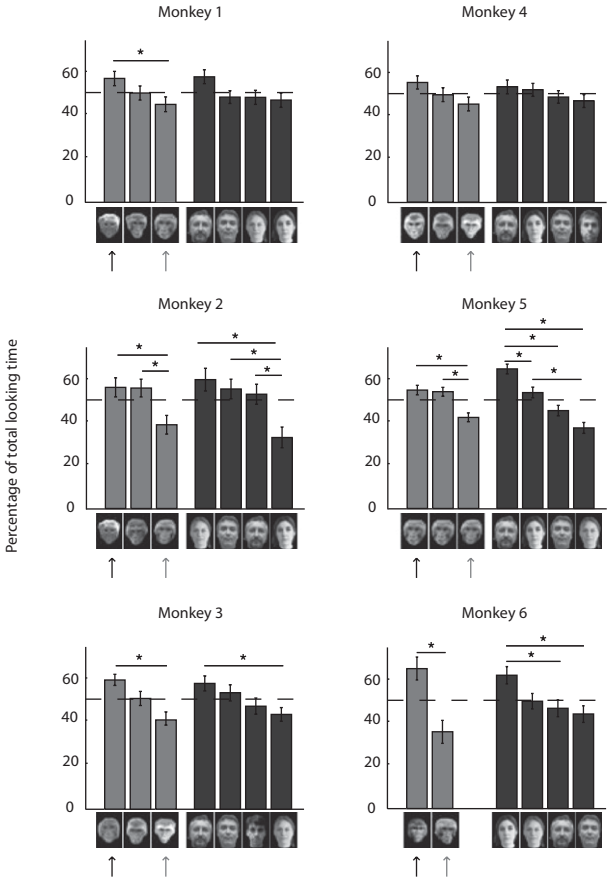

Fig. 3. Preferential looking time across individual pictures for each tested subject. Mean percentage of looking time is shown for each face picture within a pair. The dashed line indicates chance expectation. Errors bars represent SEM. Intragroup comparisons were calculated using a multiple-comparison test with a Bonferroni adjustment (*, 95% confidence intervals). Solid arrows indicate pictures of the closest monkey to each tested subject. Shaded arrows indicate pictures of the most agonistic monkey to each tested subject.

Robustness of the Results Across Individuals. To assess individual differences, the data were analyzed independently for the six tested subjects. Five of six monkeys showed significantly greater looking time for faces when preceded by matching voice relative to when they were not (Table S3 and Fig. S1*A*). Thus, the group effects described in the global analysis were not driven by isolated individuals. An ANOVA on looking times using the identity of the subject as a random-effect variable confirmed the above results, with a significant voice–face concordance factor [$F = 15.69$, df $= 1, P = 0.001$], enabling us to extrapolate the matching behavior to a population larger than the six subjects. We also tested whether cross-modal recognition of humans was relevant at the individual level. When considering separately trials with human stimuli and trials with monkey stimuli, three of six monkeys showed significantly greater looking time for faces matching the voice for monkey stimuli and five of six monkeys showed significantly greater looking time for faces matching the voice for human stimuli (Table S3 and Fig. S1*B*). This result indicates a heterogeneous behavior among individuals. Further, we tested whether this heterogeneity reflects common variations observed in a population. We also conducted two ANOVAs on looking times using the identity of the subject as a random effect but conducted separately for monkey stimuli trials [$F_M = 5.64$, df $= 1, P = 0.018$] and human stimuli trials [$F_H = 9.68$, df $= 1, P = 0.002$]. This result confirmed that the group effects were not driven by individuals' effects for trials with either monkey or human stimuli.

Examining Individual Face Photograph Preferences. Once the nature of the face–voice concordance factor was examined, we analyzed face identity, the other factor that ruled preferential looking times. Mean viewing times on face pictures were analyzed independently of voice–face concordance by summing viewing times for each face over trials preceded by the face's own voice and over trials preceded by another voice. As shown by the one-way ANOVA on the looking times performed for each tested subject, there was a strong effect of face identity factor as five of six monkeys presented significant face preferences (Table S1 and Fig. 3). To further evaluate what could account for differences in viewing times for individual faces, face pictures were arranged in decreasing order of looking time, separately for monkey and human stimuli (Fig. 3). Then, for each subject, a multiple comparison of the looking times on each face was performed, separately for trials with human and monkey stimuli (Table S1 and Fig. 3). For five of six subjects, the monkey face yielding the longest looking time was that of the monkey they were paired with or that of their neighbor, when not paired. The monkey face least looked at was that of the individual with whom they maintained the most agonistic relation, as assessed through daily observation of interactions mainly comprising threats. For three monkeys, it was the dominant monkey of the colony room and for two monkeys, it was an aggressive female. Some differences also appeared for human individual faces. It is more difficult to draw a conclusion about human individuals that would parallel the one with monkey individuals because of the difference in the nature of the interactions. Nonetheless, it appeared that the person most looked at was the researcher/experimenter in charge of the animal.

Discussion

Recently, a growing interest in determining the neural bases of social cognition has emerged (e.g., refs. 31 and 32). This new field often relies on the rhesus monkey as a model because of the close homology of its brain with the human brain. However, drawing parallels between human and rhesus monkey social cognition requires the development of behavioral approaches and testing procedures adapted to each species (33). In this study, we contributed to the goal of characterizing how rhesus monkeys process socially relevant stimuli in an experimental setting by demonstrating four key points. First, we show the existence of individual recognition in rhesus monkeys comprising at least two elements of identity (vocal and visual). Second, individual recognition extends adaptably from conspecifics to personally known humans. Third, rhesus macaques exhibit preferential bias toward pictures of certain individuals over others, likely reflecting their social interaction history. Fourth, our data indicate that rhesus macaques perceive pictures as being, if not equivalent, at least connected or related to real life stimuli.

Evidence for the Existence of Individual Cross-Modal Recognition. Overall, rhesus monkeys spontaneously looked longer at a face that matched the voice previously heard. This spontaneous cross-modal identity matching suggests that rhesus monkeys possess a cognitive representation of individuals, allowing them to link two perceptual cues. The cross-modal nature of the task is crucial. In non-cross-modal tasks, matching can be done at the perceptual level: for example, matching different views of the same face by means of common features perceived through mental rotation (34) or recognizing the different vocalizations of an individual through the presence of unchanged physical features such as formants. On the contrary, in our task, monkeys had to rely on information in memory to match one stimulus, the voice, with the other, the face. Finally, the static nature of the stimuli also affirms this conclusion. As opposed to video displays, in which faces and voices share a common dynamic that enables matching the vocal

sample to the vocal-tract movements, in a static display, their combination cannot be performed at a perceptual level but rather at a conceptual one. This result indicates that monkeys have learned through daily experience, the face–voice identity associations for familiar individuals. The relative complexity of this process explains why individual recognition has long been thought to be a uniquely human capacity, which has been reported to develop ~4 mo of age (35). However, it has recently been demonstrated in a few other species (16–19) and might be more widespread in animals (36, 37). Evidence for individual recognition by rhesus macaques raises the issue of the adaptive value of this very precise type of social recognition during interactions with conspecifics. In rhesus macaques, individual recognition might be required for some particularly sophisticated behaviors (13–15); moreover, such recognition could support third-party knowledge in those contexts wherein awareness of the peculiar relationships of others is useful in maintaining amicable relationships with a minimum of hostility and stress in complex groups (38).

Plasticity and Adaptive Potential of Individual Recognition. In the present experiment, cross-modal individual recognition extends to human stimuli, indicating that this process is not species specific but also applies to socially relevant individuals from other species who become familiar in an animal's life. It is of interest to consider the role of perceptual narrowing in the capacities observed here. Human and macaque infants possess a broad perceptual tuning at birth that later narrows for socio-ecologically relevant signals (25). Studies with monkeys have reported both narrowing for face perception (27, 5, 7) and broad face–voice dynamics perception in infants (39). Amodal attributes shared by faces and voices (tempo, collocation, or synchrony) precisely enhance association of modality-specific features (characterizing age, gender, and identity) (35, 40). In this context, then, the presence of cross-species recognition in our study could be explained by such a broad amodal face–voice perception being present at birth and persisting because of the incentive value of human individuals. This hypothesis is consistent with recent evidence in infant Japanese macaques showing the influence of exposure to conspecifics and humans on forming multisensory associations (41, 42). Whereas infant Japanese macaques raised in large groups were able to match a monkey voice to a monkey face but not a human voice to a human face, young animals with daily exposure to humans were able to match voice and face in both conditions. The latter experiments tackled coarse multisensory representation of "species," and we show here that more fine-grained multisensory representation of "identity" may also be influenced by experience with humans.

Difference in Individual Recognition for Humans and Monkeys. Surprisingly, some monkeys were equally good or better at matching human face–voice identity relative to conspecifics face–voice identity. However, we did not make direct comparisons between trials with monkey stimuli and those with human stimuli, because of the inherent inequality between human and monkey audio sets. Human speech contains more information about speaker identity than monkey coo vocalizations, making human speakers easier to recognize. Whereas monkey "coos" and human speech both contain uniform cues for individual discrimination, only human speech contains dynamic cues changing over the duration of an utterance that represent additional robust information for speaker identity (43). As a result, similarities in performance across stimulus sets could be driven by differences in the complexity of the two stimulus sets, rather than by the incentive value of one of these two groups of familiar individuals.

What Do Monkeys Perceive in the Pictures of Familiar Individuals? Previous experiments in capuchin monkeys were conducted with pictures of familiar individuals in social discrimination tasks. They revealed that 2D images of familiar conspecifics are interpreted by capuchin monkeys as representing reality (44) and therefore could be well suited to investigations of individual recognition. In studies with rhesus monkeys, pictures of familiar individuals were also used in social discrimination tasks (e.g., refs. 12 and 45). However, it was unclear whether rhesus macaques relied on pictorial features, such as prominent facial muscles that represent secondary sexual characteristics in dominant males, or on the social knowledge underpinned by them (i.e., on previously acquired knowledge of the hierarchy between the tested monkey and the one in the picture). Indeed, we know little about how rhesus monkeys perceive photographs, beyond the facts that they perceive features and colors of photographs as humans do (46), that they are highly motivated to attend to visual stimuli involving conspecifics (47), and that they find social stimuli rewarding (47). Pictures are both concrete objects with features and colors and a representation of something other than themselves, and it is unclear whether monkeys perceive a connection between the reality and the content of the picture (48, 49). Here we presented faces as cropped photographs that included head, neck, and hairs, on a black background. Compared with real faces, the stimuli were reduced along several physical dimensions, such as size, stereoscopy, and motion parallax cues. Despite the reduced informational content, monkeys extracted relevant visual identity cues, first, to match them with auditory identity recording and second, to drive preferential looking time. Thus, pictures provided at least a sufficient cue for the recall of the individual they represent, indicating that rhesus monkeys establish a connection between the reality and the content of the pictures and use it to orient their behavior. This result implies that face photographs of familiar individuals are relevant and adapted stimuli for use in behavioral or neurophysiological laboratory tests.

About Preferences of Individual Face Photographs Over Others. Rhesus monkeys displayed significant differences in their looking times across individual face pictures, favoring, among monkeys, the individual with whom they maintained the closest relation (their pair or neighboring mate) or, among humans, their main caregiver. This result suggests that experience and interactions shaped subjects' preferences. This pattern of preference contrasts with findings showing that rhesus monkeys accept lower rewards as a payoff for looking at pictures of dominant monkeys (12). However, the results of the latter study also suggest that the initial orientation toward a face and total duration of the viewing time might reflect two separate processes: Monkeys might choose to look initially at a dominant peer but spend more total time looking at other monkeys. According to the authors, orientation enables one to gain visual information that might be necessary for future social behavior, whereas total viewing duration represents a commitment into a specific relationship with the individual in the picture. This interpretation is consistent with our data, as subjects spent more time looking at individuals with whom they felt safest and avoided eye contact with monkeys with whom they usually had agonistic interactions. This behavior may also be driven by the fact that we used only affiliation calls and not threat calls, which might have led to a longer viewing time toward the agonistic monkeys.

From Voice–Face Identity Association Toward the Concept of a "Person." In humans, a complex cognitive representation binds semantic information with information of different sensory modalities. This information includes at least name, face, voice, silhouette, odor, gait, and feel of the surface of the skin (50–52). The concept of a person also implies knowledge of interaction history with a given individual and expectations about its future behavior (53). We demonstrate that rhesus macaques possess at least a multifaceted internal model of other individuals that includes face–voice association and face–photograph association for each individual. Our experiment revealed other (but not all) aspects of the concept of a person, by exhibiting strong preferences for

certain individuals over others. These data offer unique insight into the complexity of conceptual thinking about other individuals by rhesus monkeys. A simple twofold face–voice association is much less complex than the concept of a person as a whole; yet it is constitutive of this concept. This association might represent a basic building block of cognition shared across a wide range of species and may be an evolutionary precursor of the concept of a person found in humans. As such, insight into this skill is gained from investigation in a bottom–up perspective of animal and human cognition (33).

Conclusion. The present findings advance our understanding of face–voice integration in rhesus monkeys. Prior studies showed that rhesus monkeys can *i*) perform lip–speech association by matching face dynamics to corresponding vocalization, *ii*) associate the number of voices heard to the number of faces seen, and *iii*) link vocal-tract resonance to corresponding physical shape (54–56). We now demonstrate the existence of a cross-modal voice–face identity association that encompasses both familiar peers and familiar humans. Moreover, our results suggest the existence of a cognitive representation of identity based on memories of previous interactions. Consequently, behavioral studies such as this set the stage for further neurophysiological investigations of social cognition.

Materials and Methods

Subjects. Six rhesus monkeys (*Macaca mulatta*, three females, 6–11 y old) were used. Animals were socially housed in rooms of four individuals after they arrived between 2–5 y ago at the Centre de Neuroscience Cognitive. They were born and bred during 2–3 y in small rhesus macaque groups with daily exposure to both conspecifics and humans. All experimental procedures were in accordance with the local authorities (Direction Départementale des Services Vétérinaires, Lyon, France) and the European Community standards for the care and use of laboratory animals [European Community Council Directive (1986), Ministère de l'Agriculture et de la Forêt, Commission Nationale de l'Expérimentation Animale].

Stimuli. Visual and auditory test stimuli were color photographs and audio recordings of individuals familiar to the subjects (Fig. 1*A*). Familiar simian individuals were the three rhesus macaques housed in the same room as the subjects for 2–5 y before the test. The individuals were adult 6- to 16-y-old monkeys. Familiar human individuals were the two caregivers and two experimenters working with the animals on a daily basis during the same 2- to 5-y period. Each stimulus set was specific to each animal to ensure a high familiarity with individuals presented in stimuli. Front view photographs were cropped to include only the face, head, and neck and then adjusted to a common size and presented on a black background (GNU Image Manipulation Program; www.gimp.org). Humans were presented without mask and goggles as the colony rooms are equipped with windows through which animals see the experimenter and staff without their protections. There was one photograph per familiar individual, leading to four human and three monkey photographs per animal tested. Photographs shown in Figs. 1 and 3 are provided for illustrative purposes but do not correspond to the actual stimuli, which included for all animal photographs the head implants (head posts and recording chambers). Photographs used otherwise resembled the illustrative ones in all respects (color, size, and gaze direction). Auditory stimuli consisted of audio recordings of 2-s duration, each containing either a sample of coo vocalization (894 ± 300 ms) or a human speech extract (877 ± 380 ms). Human speech extracts consisted of small sentences or words in French such as "bonjour tout le monde" ("hello everybody") or "voilà" ("here it is"). Six audio samples from each familiar individual were presented, leading to 24 human voice stimuli and 18 vocalizations per animal tested. The mean sound pressure level for the duration of each audio sample was calculated. Then the 42 audio samples for each subject were normalized for this mean acoustic intensity (MATLAB; MathWorks). Visual and auditory training stimuli were fractals (Fractal Explorer; fractals.da.ru) and synthetic abstract audio samples.

Preferential Looking Paradigm. Before testing, animals first learned to complete exploration trials with abstract audio and visual stimuli (*SI Materials and Methods, Text S3* and Fig. S2). The goal of the training task was to maintain the animal's gaze centered in the middle of the screen during the audio playback and allow it to freely explore two images presented together on the screen after the audio playback, within the boundary of a virtual window comprising both stimuli. Once animals could complete 150 trials, training stimuli were replaced by test stimuli representing familiar individuals (Fig. 1 *A* and *B*). Size, positions, and timings of stimuli presentation remained unchanged between training and test task. Juice reward was maintained in the intertrial period to ensure the monkeys were motivated to complete trials. Two types of test trials were randomly interleaved: trials with stimuli representing familiar rhesus macaques and trials with stimuli representing familiar humans. In each trial a voice or vocalization playback was followed by the presentation of two known faces, only one of which matched the voice of the playback. The subject could freely explore these face photographs during 1.5 s as long as its gaze was maintained within the boundaries of a virtual window around the pair of photographs corresponding to the black area in Fig. 1*B*. The design was counterbalanced such that each face appeared equally often in the upper and lower part of the screen, and each face was equally associated with a voice-vocalization record. We limited testing to three sessions of 75 completed trials with monkey stimuli and 75 completed trials with human stimuli per subject on successive days to preserve spontaneous behavior and prevent any stereotyped behavior from occurring during sessions. Importantly at no point in the task did we encourage the animal to associate a face photograph with a voice. On the contrary, we just observed their spontaneous behavior in looking at their choice of face photograph. Animals were rewarded equally regardless of whether they looked longer at the matching face, at the nonmatching face, or at both faces equally. Further, animals were rewarded even if they chose to look at no photograph by keeping their gaze in the middle of the screen.

Procedure. For training and testing, a subject was head restrained and placed in front of, and at eye level to, an LCD screen (Dell Computers) situated at 56 cm. The subject's gaze position was monitored with an infrared eye tracker (ISCAN) with a 250-Hz sampling rate. Eye movement samples were recorded and stored using REX Software (57), which also served for real-time control of stimulus presentation, eye position monitoring, and reward delivery. The calibration procedure is detailed in *SI Materials and Methods, Text S4*.

Data Analysis. Data were analyzed with custom-written scripts and the Statistics Toolbox in Matlab (MathWorks). Looking at the upper face in the pair was defined as when the eyes' coordinates were located inside the upper part of the virtual exploration window (12° × 12°) that contained the cropped face photograph plus ≈4° of black surround. The looking time for a particular face was the amount of time the gaze was present inside of this window and was automatically computed with the Matlab routines. When the gaze was not on the upper face, it was either on the lower face or in a 12° × 2° central black area between the two faces that was not considered for the analyses. If the gaze left the exploration window, the trial was ended prematurely and was not analyzed.

To analyze whether a particular face photograph affected the results, a one-way ANOVA was performed on the looking time on each face for each subject separately. No overall ANOVA could be used because each stimulus set was specific to each animal to ensure a high familiarity with individuals presented in stimuli. To analyze the modulation of face preferences by the voice identity, for each face of each pair, the mean μ_i and variance $\sigma^2{}_i$ of D_i = (looking time spent on face$_i$ when matching the voice) − (looking time spent on face$_i$ when not matching the voice) were calculated. Before this analysis, we tested that the variances of both groups were similar with a Bartlett's test for equal variances. A *t* test was performed on the concatenation of all means μ_i to assess whether the group mean was different from zero. Further, a *t* test was also performed separately for trials with monkey stimuli and human stimuli. The change over time of the modulation of face preferences by the identity of the voice was analyzed with a sliding window of width 0.3 s moving every 0.1 s. A *t* test on the D_i calculated at each time step was performed every 0.1 s.

To analyze individual performances, the modulation of face preferences by the voice identity was analyzed for each subject as described previously, by calculating for each face of each pair the mean μ_i and variance $\sigma^2{}_i$ of D_i. The group mean and variance (μ_G, $\sigma^2{}_G$) of all of the pairs were inferred with an expectation-maximization (EM) algorithm that enables us to find their maximum-likelihood (ML) estimates via an iterative method alternating expectation (E) and maximization (M) steps. In the E-step, μ_i, and $\sigma^2{}_i$ were estimated by computing their ML estimates knowing the group mean and variance:

$$\mu_i = \frac{\sigma_G^2}{\sigma_G^2 + \widehat{\sigma_i^2}}\,\widehat{\mu_i} + \frac{\widehat{\sigma_i^2}}{\sigma_G^2 + \widehat{\sigma_i^2}}\,\mu_G, \sigma_i^2 = \frac{\sigma_G^2 \widehat{\sigma_i^2}}{\sigma_G^2 + \widehat{\sigma_i^2}}.$$

In the M-step, the group mean and variance were reevaluated, maximizing their likelihood knowing the face-pair means and variances. This maximization reduces to the following equation:

$$\mu_G = \frac{1}{n}\sum_i \mu_i, \sigma_G^2 = \frac{1}{n}\sum_i \left[\sigma_i^2 + (\mu_G - \mu_i)^2\right].$$

The new μ_G and σ^2_G parameters are then used in the next E-step. This process quickly converges toward the ML values of μ_G and σ^2_G (58). This estimate weights the contribution of face-pair results. The EM method was used to take into account the behavioral variability of the subjects for each face pair without having to classify and remove any outlier data. A *t* test was then performed to compare the group (μ_G, σ^2_G) difference with a null difference.

Finally, a three-way ANOVA was performed on the percentage of looking time with the following factors: voice–face concordance, face preference (preferred, neutral, or less preferred as assessed by a paired *t* test on each pair of faces), and identity of the tested subject as a random effect. To analyze preferences among individuals' faces, a multiple comparison with a Bonferroni adjustment of Student's' critical values was performed on the looking time for each face for each subject, for trials with human and monkey stimuli, respectively.

ACKNOWLEDGMENTS. We thank animal care staff and experimenters from the Centre de Neuroscience Cognitive who were photographed and recorded; J.-L. Charieau and F. Hérant for animal care; P. Baraduc and P. Vindras for useful discussions on statistical analyses; V. Chambon for helpful discussion on the data, and L. Parsons for help with the English editing of the manuscript. Research was supported by a Marie Curie re-integration grant and a salary grant from the Fondation pour la Recherche Médicale (to S.W.), a PhD grant from Centre National de la Recherche Scientifique and the Direction Départementale de l'Armenent under the directorship of P. Balzagette and the Fondation pour le Recherche Médicale (to J.S.), and National Institutes of Health Grant R01 HD046526 (to O.P.).

1. Campanella S, Belin P (2007) Integrating face and voice in person perception. *Trends Cogn Sci* 11:535–543.
2. Rendall D, Rodman PS, Emond RE (1996) Vocal recognition of individuals and kin in free ranging rhesus monkeys. *Anim Behav* 51:1007–1015.
3. Parr LA, Winslow JT, Hopkins WD, de Waal FBM (2000) Recognizing facial cues: Individual discrimination by chimpanzees (Pan troglodytes) and rhesus monkeys (Macaca mulatta). *J Comp Psychol* 114:47–60.
4. Gothard KM, Erickson CA, Amaral DG (2004) How do rhesus monkeys (*Macaca mulatta*) scan faces in a visual paired comparison task? *Anim Cogn* 7:25–36.
5. Gothard KM, Brooks KN, Peterson MA (2009) Multiple perceptual strategies used by macaque monkeys for face recognition. *Anim Cogn* 12:155–167.
6. Dahl CD, Logothetis NK, Hoffman KL (2007) Individuation and holistic processing of faces in rhesus monkeys. *Proc Biol Sci* 274:2069–2076.
7. Pascalis O, Bachevalier J (1998) Face recognition in primates: A cross-species study. *Behav Process* 43:87–96.
8. Waitt C, et al. (2003) Evidence from rhesus macaques suggests that male coloration plays a role in female primate mate choice. *Proc Biol Sci* 270(Suppl 2):S144–S146.
9. Waitt C, Gerald MS, Little AC, Kraiselburd E (2006) Selective attention toward female secondary sexual color in male rhesus macaques. *Am J Primatol* 68:738–744.
10. De Waal FBM, Luttrell LM (1985) The formal hierarchy of rhesus macaques: An investigation of the bared-teeth display. *Am J Primatol* 9:73–85.
11. Gouzoules H, Gouzoules S, Tomaszycki M (1998) Agonistic screams and the classification of dominance relationships: Are monkeys fuzzy logicians? *Anim Behav* 55:51–60.
12. Deaner RO, Khera AM, Platt ML (2005) Monkeys pay per view: Adaptive valuation of social images by rhesus macaques. *Curr Biol* 15:543–548.
13. Kaplan JR (1978) Fight interference and altruism in rhesus monkeys. *Am J Phys Anthropol* 49:241–250.
14. De Waal FBM, Yoshihara D (1983) Reconciliation and redirected affection in rhesus monkeys. *Behaviour* 85:224–241.
15. De Waal FBM, Luttrell LM (1988) Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethol Sociobiol* 9: 101–118.
16. Bauer HR, Philip M (1983) Facial and vocal individual recognition in the common chimpanzee. *Psychol Rec* 33:161–170.
17. Kojima S, Izumi A, Ceugniet M (2003) Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee. *Primates* 44:225–230.
18. Proops L, McComb K, Reby D (2009) Cross-modal individual recognition in domestic horses (*Equus caballus*). *Proc Natl Acad Sci USA* 106:947–951.
19. Bovet D, Deputte BL (2009) Matching vocalizations to faces of familiar conspecifics in grey-cheeked mangabeys (Lophocebus albigena). *Folia Primatol (Basel)* 80:220–232.
20. Adachi I, Kuwahata H, Fujita K (2007) Dogs recall their owner's face upon hearing the owner's voice. *Anim Cogn* 10:17–21.
21. Adachi I, Fujita K (2007) Cross-modal representation of human caretakers in squirrel monkeys. *Behav Process* 74:27–32.
22. Pascalis O, de Haan M, Nelson CA (2002) Is face processing species-specific during the first year of life? *Science* 296:1321–1323.
23. Kelly DJ, et al. (2009) Development of the other-race effect during infancy: Evidence toward universality? *J Exp Child Psychol* 104:105–114.
24. Lewkowicz DJ, Ghazanfar AA (2006) The decline of cross-species intersensory perception in human infants. *Proc Natl Acad Sci USA* 103:6771–6774.
25. Lewkowicz DJ, Ghazanfar AA (2009) The emergence of multisensory systems through perceptual narrowing. *Trends Cogn Sci* 13:470–478.
26. Pascalis O, et al. (2005) Plasticity of face processing in infancy. *Proc Natl Acad Sci USA* 102:5297–5300.
27. Sugita Y (2008) Face perception in monkeys reared with no exposure to faces. *Proc Natl Acad Sci USA* 105:394–398.
28. Rendall D, Owren MJ, Rodman PS (1998) The role of vocal tract filtering in identity cueing in rhesus monkey (Macaca mulatta) vocalizations. *J Acoust Soc Am* 103: 602–614.
29. Lachs L, Pisoni DB (2004) Crossmodal source identification in speech perception. *Ecol Psychol* 16:159–187.
30. Kamachi M, Hill H, Lander K, Vatikiotis-Bateson E (2003) "Putting the face to the voice": Matching identity across modality. *Curr Biol* 13:1709–1714.
31. Ghazanfar AA, Santos LR (2004) Primate brains in the wild: The sensory bases for social interactions. *Nat Rev Neurosci* 5:603–616.
32. Iriki A, Sakura O (2008) The neuroscience of primate intellectual evolution: Natural selection and passive and intentional niche construction. *Philos Trans R Soc Lond B Biol Sci* 363:2229–2241.
33. de Waal FB, Ferrari PF (2010) Towards a bottom-up perspective on animal and human cognition. *Trends Cogn Sci* 14:201–207.
34. Wang G, Obama S, Yamashita W, Sugihara T, Tanaka K (2005) Prior experience of rotation is not required for recognizing objects seen from different angles. *Nat Neurosci* 8:1768–1775.
35. Bahrick LE, Hernandez-Reif M, Flom R (2005) The development of infant learning about specific face-voice relations. *Dev Psychol* 41:541–552.
36. Johnston RE, Bullock TA (2001) Individual recognition by use of odors in golden hamsters: The nature of individual representations. *Anim Behav* 61:545–557.
37. Seyfarth RM, Cheney DL (2009) Seeing who we hear and hearing who we see. *Proc Natl Acad Sci USA* 106:669–670.
38. Silk JB (2007) Social components of fitness in primate groups. *Science* 317:1347–1351.
39. Zangenehpour S, Ghazanfar AA, Lewkowicz DJ, Zatorre RJ (2009) Heterochrony and cross-species intersensory matching by infant vervet monkeys. *PLoS ONE* 4:e4302.
40. Bahrick LE, Lickliter R (2000) Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev Psychol* 36:190–201.
41. Adachi I, Kuwahata H, Fujita K, Tomonaga M, Matsuzawa T (2006) Japanese macaques form a cross-modal representation of their own species in their first year of life. *Primates* 47:350–354.
42. Adachi I, Kuwahata H, Fujita K, Tomonaga M, Matsuzawa T (2009) Plasticity of ability to form cross-modal representations in infant Japanese macaques. *Dev Sci* 12: 446–452.
43. Fitch WT (2000) The evolution of speech: A comparative review. *Trends Cogn Sci* 4: 258–267.
44. Pokorny JJ, de Waal FBM (2009) Monkeys recognize the faces of group mates in photographs. *Proc Natl Acad Sci USA* 106:21539–21543.
45. Shepherd SV, Deaner RO, Platt ML (2006) Social status gates social attention in monkeys. *Curr Biol* 16:R119–R120.
46. Waitt C, Buchanan-Smith HM (2006) Perceptual considerations in the use of colored photographic and video stimuli to study nonhuman primate behavior. *Am J Primatol* 68:1054–1067.
47. Anderson JR (1998) Social stimuli and social rewards in primate learning and cognition. *Behav Process* 42:159–175.
48. Bovet D, Vauclair J (2000) Picture recognition in animals and humans. *Behav Brain Res* 109:143–165.
49. Fagot J, Thompson RKR, Parron C (2010) How to read a picture: Lessons from nonhuman primates. *Proc Natl Acad Sci USA* 107:519–520.
50. Cutting JE, Kozlowski LT (1977) Recognizing friends by their walk: Gait perception without familiarity cues. *Bull Psychon Soc* 9:353–356.
51. Porter RH, Balogh RD, Cernoch JM, Franchi C (1986) Recognition of kin through characteristic body odors. *Chem Senses* 11:389–395.
52. Kaitz M (1992) Recognition of familiar individuals by touch. *Physiol Behav* 52: 565–567.
53. Delgado MR, Frank RH, Phelps EA (2005) Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat Neurosci* 8:1611–1618.
54. Ghazanfar AA, Logothetis NK (2003) Neuroperception: Facial expressions linked to monkey calls. *Nature* 423:937–938.
55. Ghazanfar AA, et al. (2007) Vocal-tract resonances as indexical cues in rhesus monkeys. *Curr Biol* 17:425–430.
56. Jordan KE, Brannon EM, Logothetis NK, Ghazanfar AA (2005) Monkeys match the number of voices they hear to the number of faces they see. *Curr Biol* 15:1034–1038.
57. Hays AV, Richmond BJ, Optican LM (1982) A UNIX-based multiple-process system for real-time data acquisition and control. *WESCON Conference Proceedings* (National Eye Institute, Bethesda), Vol 2, pp 1–10.
58. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data with the EM algorithm. *J R Stat Soc Ser A* 39:1–38.