

Seeing who we hear and hearing who we see

Robert M. Seyfarth¹ and Dorothy L. Cheney

Departments of Psychology and Biology, University of Pennsylvania, Philadelphia, PA 19104

Imagine that you're working in your office and you hear two voices outside in the hallway. Both are familiar. You immediately picture the individuals involved. You walk out to join them and there they are, looking exactly as you'd imagined. Effortlessly and unconsciously you have just performed two actions of great interest to cognitive scientists: cross-modal perception (in this case, by using auditory information to create a visual image) and individual recognition (the identification of a specific person according to a rich, multimodal, and individually distinct set of cues, and the placement of that individual in a society of many others). An article in this issue of PNAS by Proops, McComb, and Reby (1) shows that horses do it, too, and just as routinely, without any special training. The result, although not surprising, is nonetheless the first clear demonstration that a non-human animal recognizes members of its own species across sensory modalities. It raises intriguing questions about the origins of conceptual knowledge and the extent to which brain mechanisms in many species—birds, mammals, as well as humans—are essentially multisensory.

Individual Recognition

Individual recognition, based on auditory, visual, or olfactory cues, is widespread in animals (2). Its adaptive value is clear. Recognizing others as distinct individuals allows an animal to identify and remember those with whom it may have subtly different competitive or cooperative relations, and to place them in the appropriate social context. Experiments on monkeys, for example, suggest that listeners recognize others individually by voice (3) and make use of this information when responding to calls according to an individual's current mating status (4), unique dominance rank (5), membership in a particular kin group (6), or rank and kinship combined (7). When a female baboon is separated from her offspring and hears the offspring's call, she looks toward the sound of the vocalization (8); when female baboons and vervet monkeys hear unrelated juveniles call, they look toward the juvenile's mother (6, 9).

Individual recognition is most often documented in the auditory mode, through playback experiments. In the studies cited above, however (and many others like them), it is difficult to escape the impression that animals are engaged

in cross-modal or even multimodal processing. A baboon who looks toward the source of the sound when she hears her offspring's call acts as if the sound has created an expectation of what she will see if she looks in that direction. Humans, of course, do this routinely, integrating information about faces and voices to form the rich, multimodal percept of a *person* (10).

The first evidence that animals might integrate multiple cues to form a representation of an individual came from work by Johnston and colleagues on hamsters (11). Golden hamsters have at

Individual recognition, based on auditory, visual, or olfactory cues, is widespread in animals.

least five different odors that are individually distinctive. In a typical experiment, a male familiar with females A and B was exposed (and became habituated to) the vaginal secretions of female A. He was then tested with either A's or B's flank secretions. Males tested with A's flank secretions showed little response (across-odor habituation); however, males tested with B's flank secretions responded strongly. The authors concluded that "when a male was habituated to one odor he was also becoming habituated to the integrated representation of that individual" and was therefore not surprised to encounter a different odor from the same animal. Hamsters, they suggest, have an integrated, multi-odor memory of other individuals. Recent experiments indicate that direct physical contact with an individual—not just exposure to its odors—is necessary for such memories to develop (12).

But what about the representation of individuals across sensory modalities? Laboratory studies have shown that dogs (13) and squirrel monkeys (14) associate the faces and voices of their caretakers, but until now there has, somewhat surprisingly, been no test for cross-modal recognition of conspecifics. In the current study, Proops *et al.* (1) began by observing horses in two captive herds of ≈ 30 individuals. After re-

coding six whinnies from four different individuals, they used a violation-of-expectation paradigm to test for cross-modal individual recognition. Each subject ($n = 24$) saw one of two herd companions walk past him and disappear behind a barrier. After a delay, the subject heard from behind the barrier a whinny from either the same or a different individual. The investigators predicted that, if subjects were capable of forming cross-modal or multimodal representations of specific individuals, then the sight of individual X disappearing behind a barrier followed by the sound of X's whinny would entail no surprise, whereas the sight of X disappearing followed by Y's whinny would violate their expectations. And this is what they found: subjects responded more quickly (by looking toward the speaker), looked longer, and looked more often in the incongruent than the congruent condition.

Underlying Mechanisms

As the first study to demonstrate cross-modal integration of information about identity in animals, the experiment by Proops and colleagues (1) is likely to stimulate similar tests on many other species. Indeed, ethologists have some way to go before they catch up to neurophysiologists, who have been actively investigating sensory integration in the brain in the past few years. For example, both Poremba (15) and Gil da Costa *et al.* (16) found that, when rhesus macaques hear one of their own species vocalizations, they exhibit neural activity not only in areas associated with auditory processing, but also in higher-order visual areas, including the superior temporal sulcus (STS), an area that is known to be involved in recognizing talker identity in both humans (17) and monkeys (18). Auditory and visual areas also have extensive anatomical connections (19). Ghazanfar *et al.* (20) studied cross-modal integration by using the coos and grunts of rhesus macaques. They found clear evidence that cells in certain areas of the auditory cortex were more responsive to bimodal (visual and

Author contributions: R.M.S. and D.L.C. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 947.

¹To whom correspondence should be addressed. E-mail: seyfarth@psych.upenn.edu.

© 2009 by The National Academy of Sciences of the USA

auditory) than to unimodal presentation of calls. Although significant integration of visual and auditory information occurred in trials with both vocalizations, the effect of cross-modal presentation was greater with grunts than with coos. The authors speculate that this may have occurred because, under natural conditions, grunts are usually directed toward a specific individual in dyadic interactions, whereas coos tend to broadcast generally to the group at large. The greater cross-modal integration in the processing of grunts may therefore have arisen because, in contrast to listeners who hear a coo, listeners who hear a grunt must immediately determine whether or not the call is directed at them—and this, in turn, may depend crucially on memories of whom the listener has interacted with in the immediate past. Field experiments suggest that the memory of recent interactions with particular individuals determines whether baboons judge a vocalization to be directed at them or at someone else (21).

What neural mechanisms underlie cross-modal integration? According to a traditional view, multisensory integration takes place only after extensive unisensory processing has occurred (22).

Multimodal (or amodal) integration is a higher-order process that occurs in different areas from unimodal sensory processing, and different species may or may not be capable of multisensory integration. Perhaps as a result, different species may or may not form what in humans constitutes an integrated, multimodal (or amodal) conceptual system.

An alternative view argues that, although different sensory systems can operate on their own, sensory integration is rapid, pervasive, and widely distributed across species. The result is a distributed circuit of modality-specific subsystems, linked together to form a multimodal percept. Or, as Barsalou (23) describes the processing of calls by monkeys, “the auditory system processes the call, the visual system processes the faces and bodies of conspecifics, along with their expressions and actions, and the affective system processes emotional responses. Association areas capture these activations ... storing them for later representational use. When subsequent calls are encoded, they reactivate the auditory component ... which in turn activates the remaining components in other modalities. Thus the distributed property circuit that processed the

original situation later represents it conceptually.”

A third view argues that many neurons are multisensory, able to respond to stimuli in either the visual or the auditory domain (for example), and capable of integrating sensory information at the level of a single neuron as long as the two sorts of information are congruent. As a result, “much, if not all, of neocortex is multisensory” (24). By this account, perceptual development does not occur in one sensory modality at a time but is integrated from the start (25).

Whatever the underlying mechanism, it is now clear that individual recognition is pervasive throughout the animal kingdom. The experiment by Proops *et al.* (1) further suggests that cross-modal processing of individual identity is equally widespread and that a rich, multimodal or amodal representation underlies animals’ recognition of others. These results suggest that the ability to form an integrated, multisensory representation of specific individuals (a kind of concept) has a long evolutionary history. Perhaps the earliest concept—whenever it appeared—was a social one: what in our species we call the concept of a person.

- Proops L, McComb K, Reby D (2008) Cross-modal individual recognition in domestic horses (*Equus caballus*). *Proc Natl Acad Sci USA* 106:947–951.
- Tibbetts EA, Dale J (2007) Individual recognition: It is good to be different. *Trends Ecol Evol* 22:529–537.
- Rendall D, Rodman PS, Emond RE (1996) Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim Behav* 51:1007–1015.
- Crockford C, Wittig R, Seyfarth RM, Cheney DL (2007) Baboons eavesdrop to deduce mating opportunities. *Anim Behav* 73:885–890.
- Cheney DL, Seyfarth RM, Silk JB (1995) The responses of female baboons to anomalous social interactions: Evidence for causal reasoning? *J Comp Psychol* 109:134–141.
- Cheney DL, Seyfarth RM (1999) Recognition of other individuals’ social relationships by female baboons. *Anim Behav* 58:67–75.
- Bergman T, Beehner J, Seyfarth RM, Cheney DL (2003) Hierarchical classification by rank and kinship in female baboons. *Science* 302:1234–1236.
- Rendall D, Cheney DL, Seyfarth RM (2000) Proximate factors mediating ‘contact’ calls in adult female baboons and their infants. *J Comp Psychol* 114:36–46.
- Cheney DL, Seyfarth RM (1980) Vocal recognition in free-ranging vervet monkeys. *Anim Behav* 28:362–367.
- Campanella S, Belin P (2007) Integrating face and voice in person perception. *Trends Cog Sci* 11:535–543.
- Johnston RE, Bullock TA (2001) Individual recognition by use of odors in golden hamsters: The nature of individual representations. *Anim Behav* 61:545–557.
- Johnston RE, Peng A (2008) Memory for individuals: Hamsters (*Mesocricetus auratus*) require contact to develop multicomponent representations (concepts) of others. *J Comp Psychol* 122:121–131.
- Adachi I, Kuwahata H, Fujita K (2007) Dogs recall their owner’s face upon hearing the owner’s voice. *Anim Cogn* 10:17–21.
- Adachi I, Fujita K (2007) Cross-modal representation of human caretakers in squirrel monkeys. *Behav Proc* 74:27–32.
- Poremba A, *et al.* (2004) Species-specific calls evoke asymmetric activity in the monkey’s temporal poles. *Nature* 427:448–451.
- Gil da Costa R, *et al.* (2004) Toward an evolutionary perspective on conceptual representation: Species-specific calls activate visual and affective processing systems in the macaque. *Proc Natl Acad Sci USA* 101:17516–17521.
- Belin P, Zatorre RJ (2003) Adaptation to speaker’s voice in right anterior temporal-lobe. *NeuroReport* 14:2105–2109.
- Petkov CI, *et al.* (2008) A voice region in the monkey brain. *Nat Neurosci* 11:367–374.
- Cappe C, Barone P (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *Eur J Neurosci* 22:2886–2902.
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis N (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25:5004–5012.
- Engh AL, Hoffmeier RR, Cheney DL, Seyfarth RM (2006) Who, me? Can baboons infer the target of vocalizations? *Anim Behav* 71:381–387.
- Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1:1–47.
- Barsalou L (2005) Continuity of the conceptual system across species. *Trends Cog Sci* 9:309–311.
- Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cog Sci* 10:278–285.
- Lewkowicz DJ, Ghazanfar AA (2006) The decline of cross-species intersensory perception in human infants. *Proc Natl Acad Sci USA* 103:6771–6774.