

# Animal awareness: The (un)binding of multisensory cues in decision making by animals

Ron Hoy\*

Department of Neurobiology and Behavior, 215 Mudd Hall, Cornell University, Ithaca, NY 14850; and Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720

Each day, we perceive the world unfolding before us, and we never give a thought to having to integrate the separate sights and sounds of everyday life. They are effortlessly composed in our brain as the successive moments of our conscious lives. Cognitive neuroscientists know that the neurosensory mechanisms responsible for our seamless perceptions are astonishing (1). However, the perceptual unity of our world can “break” from neurological disorders (2) and, less dramatically, when we experience sensory illusions (3, 4). How our brain keeps the many trains of sensory information “on track and in time” in perceptual space/time is fascinating. However, humans are not the only animals on Earth that confront a myriad of sights and sounds and have to make adaptive sense of them. Ethologists and behavioral ecologists have shown that animals interacting in small groups or large societies constantly make behavioral decisions, for example, whether to court and then mate or reject, or whether to challenge and then fight or flee, decisions that are deeply consequential for an individual due to the forces of natural and sexual selection (5). Such interactions are laden with sensory cues and information from multiple modalities that the individual must translate into actions that make adaptive sense.

Understanding what composes the perceptual world of animals is more challenging than self/human studies, because investigating humans is just a matter of a subject’s ability to report “what happens” in plain speech, often as her/his brain is being scanned (6). The perceptual world of animals must be inferred from their reactions to experimental manipulation, which are mostly measured by their movements. Thus, the study by Narins *et al.* in this issue of PNAS (7) will be welcome to ethologists, comparative psychologists, and cognitive neuroscientists. Narins *et al.* designed instrumentally ingenious experiments to dissect the aggressive/territorial behavior of poison-dart frogs (*Epipedobates femoralis*) and framed their findings in the context of cognitive neuroscience and human psychophysics. There is an important con-

tribution to discussions about “animal awareness” (8).

Narins *et al.* (7) worked in Guyana, in the Amazonian rain forest. There, male poison-dart frogs vigorously hold and defend territories against other conspecific males. The territorial males produce vocal and repetitive advertisement calls that attract females for mating (9). The diurnally active males can be seen as well as heard in the act of calling. A singing frog is easily recognized on sight/site by his conspicuously inflated and pulsating vocal sac, which when fully inflated is nearly half the size of the frog itself. The species-specific call is easily recognized. It consists of four loud high-pitched notes, delivered in *glissando*. When a male intrudes upon the territory of a calling resident male, he is quickly approached and attacked by the resident (9). Narins *et al.* ingeniously used what they called an “electromechanical model frog,” herein called “robofrog.” Robofrog is an accurately sculpted and painted silicon model of a male poison-dart frog, posed in singing position next to a small wooden log, wired with a small hidden loudspeaker from which prerecorded calls were broadcast. Full functionality was conferred on the model frog by giving it a “vocal sac,” consisting of an ultrathin flexible latex membrane pouch, which could be inflated and pulsed in the size and shape of a real vocal sac. Broadcasts from the loudspeaker, as well as inflation of the vocal sac, were remote controlled by investigators who observed all encounters on live-action video. Narins *et al.* first identified the territorial boundaries of frogs on their field site, then placed their robofrog within a territory and recorded the reactions of the resident frog.

I take pains to describe the experimental setup, because the ingenuity of the robotic device and ability to videotape behavior under entirely natural conditions were essential to revealing the full play of behavioral acts that unfolded in the rain forest, which might not be as robust if the experiment were performed in the sterile conditions of a laboratory. Narins *et al.* (7) discovered that broadcasting the calling sound of another male from the territory of a territorial male was sufficient to attract

the territory-holding male to the loudspeaker. Once the territorial male was attracted to the sound of the call, the sight of the silicon robofrog motivated him to touch and explore the model. However, exploration did not escalate into a full-blown aggressive attack unless the model was “calling,” simulated by broadcasting song through its nearby loudspeaker and pulsating its inflated latex “vocal sac.” Thus, Narins *et al.* fractionated the aggressive behavior of a poison-dart frog by separately dissociating the model’s vocal from visual signals: sound alone was sufficient to elicit attraction and exploration; sight alone of an accurately painted model frog would also elicit exploration and touching, but only the combination of the sight of a calling frog with the sound of its call elicits full aggression, an attack upon the model. These observations clearly illustrate the synergistic and modulatory roles of separate sensory modalities in a naturally multimodal signal. They also show the importance of ecological context in a behavior as complex as territorial defense. There is a growing appreciation of multimodal signaling in the communication of animals, ranging from invertebrates (10, 11) to vertebrates (12, 13).

In the present study, Narins *et al.* (7) went further in dissociating acoustic from visual cues to assess temporal and spatial factors that influence perceptual coherence in multimodal displays. Again, aggressive behavior served as the behavioral assay. First, time delays were introduced between the visual (inflated pulsating vocal sac) and acoustic (onset and duration of advertisement call) signs to explore the ability of the frogs to (dis)integrate time cues between modalities. Second, the experimenters dissociated the visual sign of the “singing” robofrog from the actual location of the sound source by placing a second remote loudspeaker at various distances away from the model. This permitted Narins *et al.* to quantitatively assess the ability of frogs to (dis)integrate spatial/distance cues between modalities.

See companion article on page 2425.

\*E-mail: rrrh3@cornell.edu.

© 2005 by The National Academy of Sciences of the USA

To investigate the temporal integration of perception, Narins *et al.* (7) systematically desynchronized the amplitude-modulated visual sign of the calling act, the inflation–deflation cycles of the pulsating “vocal sac” of the robofrog, from the auditory sign, the amplitude-modulated features of the call itself. The degree of desynchronization ranged from partial synchrony to complete decoupling. Live territorial male frogs were attracted to the immediate locale of the call regardless of whether the calls were synchronized with vocal sac movements of the robofrog. The stereotyped species-specific calls function as long-distance attraction signals; visual cues play no role in attraction. However, once a male has been attracted to the site/sight of the acoustically broadcasting robofrog, the amount of time it remained in the vicinity depends largely on whether the model’s visual and auditory cues are synchronized and/or overlapped; when the two cues were completely decoupled, the frog left the site much sooner, he seemingly “losing interest.” However, the male was provoked to attack the model if the visual stimulus was decoupled from the auditory stimulus by less than a half second; longer desynchronized intervals were not provocative. Thus, the sound of a calling male compels investigation, and the sight of noncalling male is not apparently threatening enough to evoke aggressive behavior, unless the two stimuli coincide within a half-second time window. The curious male escalated to full aggression only by the sight and sound of a rival male in the complete act of calling.

To test for spatial integration, Narins *et al.* (7) added an external loudspeaker to their experiment in addition to the speaker built into the model frog/log setup. This allowed them to separate the actual acoustic location of the call from the visual stimulus of the robofrog exercising its vocal sac, as though calling. By systematically varying the distance between the robofrog and the external speaker (displacements) and observing the effect on the aggressive behavior of the resident male frog, the investigators determined the limits of a male to spa-

tially integrate discordant sensory cues. As before, the resident frog was attracted to sound and investigated both the external loudspeaker and the robofrog. However, a frog’s behavioral reaction to the model depended on the displacement distance. For small spatial displacements between the model and the external speaker (defined as 2–12 cm), many physical contacts and attacks (>75%) were made upon the model. However, for large displacements (25–50 cm), the model was touched or attacked in only 25% of trials. Moreover, the amount of time the males spent in the vicinity of the robofrog was greater when the displacements were small. Apparently, to an inquisitive male frog, the more coherent its perception of a full-blown singing frog on its territory, the more compelling, and the more likely it is to attack.

Narins *et al.* (7) interpret their findings in terms of a perceptual “binding” problem. In cognitive neuroscience, the usual framing of perceptual binding is within the context of a single modality, usually mammalian vision (14, 15). Multiple visual streams diverge from the primary visual cortex, V1, and extract different features of the visual scene: movement and stereopsis, color and texture, and specially “labeled” features such as faces (15). In the brain, the confluence of disparate visual processing streams to produce a coherent and unitary visual percept has yet to be fully understood; however it is achieved, binding of multiple information streams must occur within the spatial and temporal constraints of the brain. Narins *et al.* frame their results as a multimodal binding problem, but the implications for perceptual coherence are similar. They manipulated the temporal and spatial constraints of binding visual to auditory cues and were able to “break” the coherence of perception (where the visual location of a vocalizing frog no longer “matched” the location of the sound source) in both space and time. Such inferences about the coherence of multimodal percepts in animals are hard to make, and rarely can they be tested in as direct a manner as demonstrated here.

Narins *et al.* (7) also relate their findings to human psychoacoustics by bridging them to “a number of human studies (that) have shown that visual cues can modulate the apparent location of auditory cues. This is clearly seen in the ‘spatial ventriloquist effect’ . . .” They also cite recent studies in humans that demonstrate temporal spatial ventriloquism effects. These are illusions more familiar, perhaps, to readers “of a certain age,” for whom television performances of vaudevillians performed with dummies (think Edgar Bergen and Charlie McCarthy) are memorable. Bergen “sold” the illusion that his dummy, Charlie, was speaking by moving the dummy’s lips as Bergen himself spoke Charlie’s dialogue; critical to the illusion, Bergen did not move his own lips. Accomplished ventriloquists were said to be able to “throw their voice,” a testament to how compelling the illusion could be. Linguistic studies of human speech indicate that where sound and visual cues appear to be in conflict, vision tends to dominate (16). Few comparable animal studies exist, hence the Narins *et al.* study is a welcome one. The poison-dart frogs also deal with a disparity of spatial cues in a way consistent with visual dominance, but only up to certain spatial limits, beyond which the frogs’ attention is dominated or “captured” by the auditory stimulus [figure 3 and table 3 of Narins *et al.* (7)].

This study by Narins *et al.* (7) is rich in possibilities for inferring important perceptual mechanisms, such as selective attention, sensory binding, sensory dominance, and multimodal interactions, all of which are lively issues in human cognition and perception studies but are just emerging in comparative animal studies. Narins *et al.* show once again how important it is to perform experimental tests of animal perception and cognition within natural settings, echoing the precepts of the late James J. Gibson (17), who in his own studies of human perception pointed out, over a half century ago, the importance of the “ecological validity” of experimental settings and the design of salient stimulus situations. The present study makes very clear that this is no less true for understanding the perceptual world of animals.

- Bolhuis, J., ed. (2000) *Brain, Perception, Memory* (Oxford Univ. Press, New York).
- Sacks, O. (1985) *The Man Who Mistook His Wife for a Hat and Other Clinical Tales* (Simon & Schuster, New York).
- Zihl, J., Von Cramon, D. & Mai, N. (1983) *Brain* **106**, 313–340.
- Purves, D. & Lotto, B. (2003) *Why We See What We Do: An Empirical Theory of Vision* (Sinauer, Sunderland, MA).
- Krebs, J. R. & Davies, N. B. (1993) *An Introduction to Behavioural Ecology* (Blackwell, London).
- Cabeza, R. & Kingstone, A. (2001) *Handbook of Functional Neuroimaging of Cognition* (MIT Press, Cambridge, MA).
- Narins, P. M., Grabul, D. S., Soma, K. K., Gaucher, P. & Hödl, W. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 2425–2429.
- Griffin, D. (2001) *Animal Minds* (Univ. of Chicago Press, Chicago).
- Narins, P. M., Hödl, W. & Grabul, D. S. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 577–580.
- Hebets, E. A. & Papaj, D. R. (2005) *Behav. Ecol. Sociobiol.* **57**, 197–294.
- Elias, D., Hebets, E., Hoy, R. & Mason, A. (2005) *Anim. Behav.*, in press.
- Partan, S. & Marler, P. (1999) *Science* **283**, 1272–1273.
- Patricelli, G. L., Uy, J. A. C., Walsh, G. & Borgia, G. (2002) *Nature* **415**, 279–280.
- Engel, A. K. & Singer, W. (2001) *Trends Cognit. Sci.* **5**, 16–25.
- Purves, D. (2002) *Neuroscience* (Sinauer, Sunderland, MA).
- McGurk, H. & MacDonald, J. (1976) *Nature* **264**, 746–748.
- Gibson, J. J. (1950) *The Perception of the Visual World* (Houghton Mifflin, Boston).