Knowing What to Imitate and Knowing When You Succeed

Brian Scassellati
MIT Artificial Intelligence Lab
545 Technology Square
Cambridge MA 02139, USA
scaz@ai.mit.edu
http://www.ai.mit.edu/people/scaz/

Knowing What to Imitate and Knowing When You Succeed

Imitation is an extremely powerful mechanism for social learning. Our group has been attempting to build robots that flexibly acquire skills by imitating a human instructor. Our robotic platforms have human-like appearance and mechanical structure both to allow the human instructor to have a more natural interaction with the robot and to simplify the problem of mapping actions between bodies. Our current research has focused primarily on the construction of a variety of social skills which are functional precursors of imitation, including joint reference, facial expression and gesture identification, and habituation and attentional systems.

Imitation of human actions is a complex process that involves many perceptual, cognitive, and motor capabilities. In this abstract, I would like to focus on two issues that we have been studying while trying to build systems that use imitation as a social learning tool. First, how do you know what aspects of your sensory experience are important to a successful imitation? Second, how do you know when you have been successful? Before addressing these two questions, I will briefly describe the robotic platforms that we have been constructing.

1 Robotic Platforms

Our work with imitation has focused on two platforms: an upper-torso humanoid robot called Cog and an active vision system enhanced with facial features called Kismet (see Figure 1). Both of these robots have been constructed in part to investigate how to build intelligent robotic systems by following a developmental progression of skills similar to that observed in human development (Brooks & Stein 1994, Brooks, Breazeal (Ferrell), Irie, Kemp, Marjanović, Scassellati & Williamson 1998). In the past two years, a basic repertoire of perceptual capabilities and sensory-motor skills have been implemented on the robot (see Brooks et al. (1998) for a review).

Our humanoid robot approximates a human being from the waist up with twenty-one degrees-of-freedom (DOF) and a variety of sensory systems. The physical structure of the robot, with movable torso, arms, neck and eyes gives it human-like motion, while the sensory systems (visual, auditory, vestibular, and proprioceptive) provide rich information about the robot and its immediate environment. These together present many opportunities for interaction between the robot and humans.

The robot Kismet is based around the same active vision system used on Cog. In addition to the three degrees-of-freedom in the eyes, Kismet also has one degree-of-freedom in the neck and eleven degrees-of-freedom in facial expressions, including eyebrows (each with two degrees-of-freedom: lift and arch), ears (each with two degrees-of-freedom: lift and rotate), eyelids (each with one degree of freedom: open/close), and a mouth (with one degree of freedom: open/close). The robot is able to show expressions analogous to anger, fatigue, fear, disgust, excitement, happiness, interest, sadness, and surprise in response to perceptual stimuli (Breazeal & Scassellati 1998).

By focusing on robotic platforms that are anthropomorphic, we simplify the problems of social interaction in three ways. First, it allows for a simple and natural means of interaction. People already know how to provide the robot with appropriate feedback, how to attract its attention, and what capabilities it might possess. Second, the responses of the robot can be easily identified and

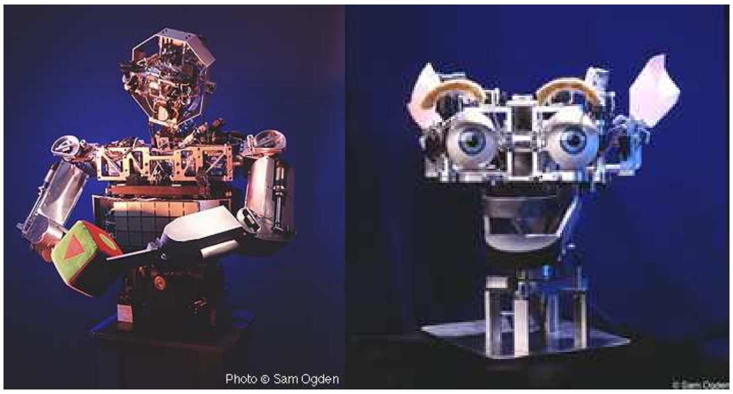

Figure 1: Two robotic platforms used in our imitation research. At left, the upper-torso humanoid robot called Cog. At right, an active vision system with facial expressions called Kismet.

interpreted by a naive observer. Third, by having a similar body structure, the problem of mapping observed actions onto the robot's own body is simplified.

We are interested in building systems that can imitate a human for a variety of reasons. From an engineering perspective, imitation provides a simple and flexible way of allowing the robot to acquire new skills without any additional programming. From a computer science perspective, imitation provides a means for biasing learning and interaction. From a developmental psychology perspective, building systems that learn through imitation allows us to investigate the minimal set of competencies necessary for social learning.

2 How do you know what to imitate?

One of the most difficult problems in complex robotic systems is determining which of the incoming sensory signals are relevant to the current task. If the robot's current goal is imitation, how does it determine which aspects of its sensory environment are relevant? Even once the robot has identified the human instructor, how does it determine which of the instructor's actions are relevant to the task and which are circumstantial?

Our research on this issue has focused on two areas: recognizing inherent saliency in objects and recognizing objects and actions that are salient as a result of the attentional state of the instructor. To recognize inherent object saliency, we have been constructing attentional and perceptual systems that combines information on visual motion, innate perceptual classifiers such as face detectors, color saliency, depth segmentation, and auditory information with a habituation mechanism and a motivational and emotional model. This attentional system will allow the robot to selectively direct computational resources and exploratory behaviors toward objects in the environment that have inherent saliency. While these attentional systems provide context-dependent saliency information, we also utilize the observed attentional states of the human instructor as a means of determining which actions and objects are relevant. We already have perceptual systems that allow us to detect faces, move the eyes to the detected face, and obtain a high-resolution image of the instructor's eyes (Scassellati 1998). We are currently working on utilizing information on the location of the pupil, the angle of gaze, the orientation of the head, and body posture to determine the object of the instructor's attention. This emphasis on joint reference is part of a larger project to build a "theory of mind" for the robot, which would allow it to attribute beliefs, desires, and intentions to the instructor (Scassellati 1999).

3 How do you know when you have been successful?

A second problem for robotic systems is success criteria. An autonomous robot must have some means of estimating how successful a previous action has been in order to further refine that behavior and to plan successive actions. This is often a difficult perceptual problem. In the case of imitation, the difficulty of obtaining a success criterion can be simplified by exploiting the natural structure of social interactions. Imitation is often an iterative process. The instructor demonstrates, the student performs, and then the instructor demonstrates again, often exaggerating or focusing on aspects of the task that the student failed to perform successfully. By responding to the same types of social cues that initially allowed the robot to understand and identify which salient aspects of the scene to imitate, the robot can incrementally refine its approximation of the actions of the instructor.

With our robots, we plan on using joint reference as a cue for iterative refinement. We have also planned an implementation of an auditory system capable of detecting prosody (including pitch, tempo, and tone of voice) as a secondary signal for obtaining feedback on which actions have been successfully executed and which have not. We are also currently engaged in building a facial expression recognition system which would allow the robot to obtain feedback directly from the facial expressions of the instructor. We have already successfully deployed a system that classifies the head motion of the instructor and allows the robot to obtain feedback from yes and no head nods.

4 Imitation behaviors

Many of the precursor skills that we have described above have already been implemented on our robots, including face detection, orientation behaviors, a basic attentional system, eye finding, head motion tracking, motion detection, and color saliency estimation. Many more of these skills are currently under development (especially an advanced attentional system, gaze direction identification, and facial gesture recognition). Using the face tracking system with some fixed motor patterns, our robots are capable of imitating yes and no head nods (see Figure 2). In the near future, we hope to enable Kismet to imitate facial expressions similar to those observed in human infants by Meltzoff & Moore (1989).

5 Acknowledgments

Support for this project is provided in part by an ONR/ARPA Vision MURI Grant (No. N00014-95-1-0600). The author wishes to thank Rod Brooks and the other members of the Cog project for their support.

References

Breazeal, C. & Scassellati, B. (1998), 'Infant-like Social Interactions between a Robot and a Human Caretaker', *Adaptive Behavior*. In submission.

Brooks, R. A. & Stein, L. A. (1994), 'Building brains for bodies', *Autonomous Robots* **1**(1), 7–25.

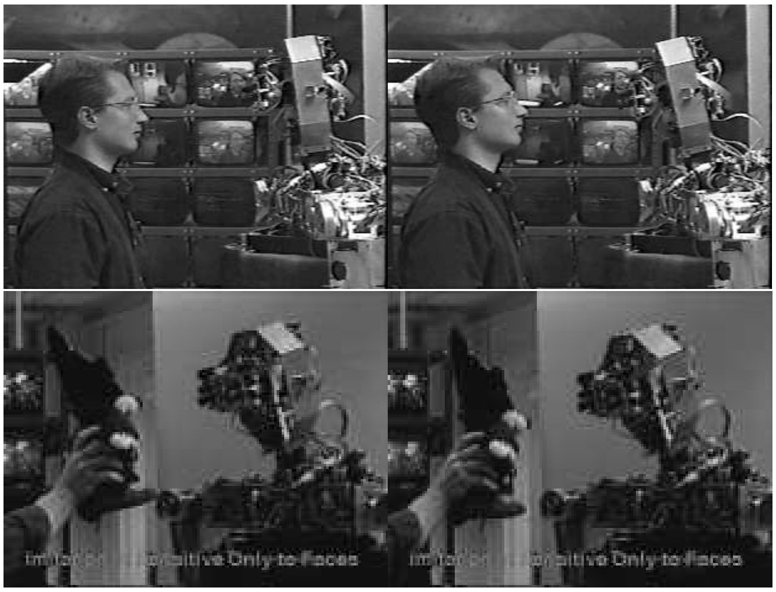

Figure 2: Images captured from a videotape of the robot imitating head nods. The upper two images show the robot imitating head nods from a human caretaker. The output of the face detector is used to drive fixed yes/no nodding responses in the robot. The face detector also picks out the face from stuffed animals, and will also mimic their actions. The original video clips are available at `http://www.ai.mit.edu/projects/cog/`.

Brooks, R. A., Breazeal (Ferrell), C., Irie, R., Kemp, C. C., Marjanović, M., Scassellati, B. & Williamson, M. M. (1998), Alternative Essences of Intelligence, *in* 'Proceedings of the American Association of Artificial Intelligence (AAAI-98)'.

Meltzoff, A. N. & Moore, M. K. (1989), 'Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms', *Developmental Psychology* **25**, 954–962.

Scassellati, B. (1998), Finding Eyes and Faces with a Foveated Vision System, *in* 'Proceedings of the American Association of Artificial Intelligence (AAAI-98)'.

Scassellati, B. (1999), Imitation and Mechanisms of Joint Attention: A Developmental Structure for Building Social Skills on a Humanoid Robot, *in* C. L. Nehaniv, ed., 'Computation for Metaphors, Analogy and Agents', Vol. 1562 of *Springer Lecture Notes in Artificial Intelligence*, Springer-Verlag.