Emergence of imitation mediated by objects

Hideki Kozima **Cocoro Nakagawa** **Hiroyuki Yano**

Communications Research Laboratory
Hikaridai, Seika-cho, Soraku-gun,
Kyoto 619-0289, Japan
{xkozima, kokoro-n, yano}@crl.go.jp

Abstract

This paper describes our model of the emergence of a mirror system for imitative learning. The mirror system is an intermodal mapping between someone's action (as seen) and one's own action (to execute). It is widely assumed that the mapping is a geometric transformation between bodies in the visual and motor spaces; however, it is still unclear if the mapping is innate or, if not, how it is acquired. We claim here that the mapping is *not* a geometric transformation between bodies but a functional correspondence between someone's action on an object for producing an effect *and* one's own action on the same or similar object for producing a similar effect, which is learnable because both tend to utilize the object's affordance in a similar way.

1. Introduction

Imitation of others' behavior is the learning method unique to *Homo sapiens*, i.e. our species (Tomasello 1999). This strategy requires a *mirror system*, that is a mapping between someone's behavior (how *he/she* feels and acts) and the self's behavior (how *I* feel and act). The mapping between how *he/she* perceives and how *I* perceive the object of interest is learnable through *local enhancement* (established by joint attention, for instance). However, the mapping between how *he/she* acts and how *I* act is certainly enigmatic, since *his/her* action is mainly visually observed and *my* own action is represented by a motor program or proprioception attached to it.

One of the remarkable advances in the study of mirror systems is the discovery of *mirror neurons* in a macaque's premotor cortex (Rizzolatti 1998). Each mirror neuron activates *both* when the macaque observes someone else perform a specific manual task *and* when the macaque performs the same task. One might say that imitative learning is straight forward, if the macaque is innately equipped with the mirror neurons. This is, however, hard to imagine, because non-human primates, especially macaques, do not engage in imitative learning (Tomasello 1999), and because such an innate mapping could not evolve rapidly enough to accommodate newly invented objects in the environment.

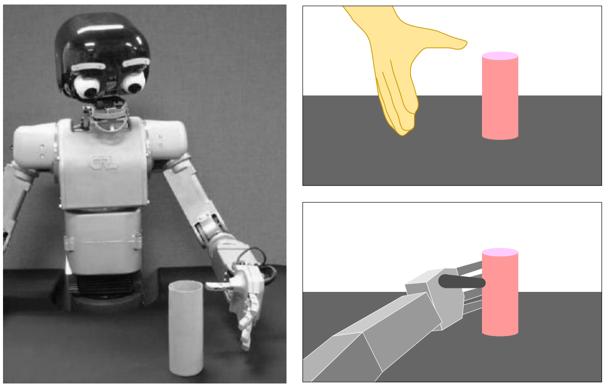

Fig. 1 Self-other mapping mediated by objects.

We claim that the mirror system is *constructed* through the interaction with the physical and social environment. A human infant can learn the correlation between how he/she responds to an object to make an effect on it and how his/her caregiver responds to the same or similar object for a similar effect, because both have functionally similar bodies and therefore respond to the object's affordance in a similar way and evaluate the produced effect in a similar way. Learning the correlation of responses to objects of interest, the infant develops a mapping between someone else's action and the infant's own action on various kinds of objects (Fig. 1). This functional mirroring would be the core mechanism not only for imitative learning but also for empathetic understanding of others' mind, both of which are specific to our species.

2. Functional mirroring

We define an action a as a body movement that is applied to an object o and produces an effect e:

$$o \xrightarrow{a} e.$$

The object o may have a bunch of affordances, each of which is a tangible feature that facilitates a relevant action on o; in other words, the action a picks up one of the affordances that is relevant to producing the effect e. The effect e has (a) a subjective aspect (intangible for others) and (b) an objective aspect (observable by others); for example, grasping a mug gives a subjective effect of haptic and proprioceptive sensation of holding the mug and an objective effect of getting the mug to lift off from the table. For the time being, we will consider only the observable effects on the object.

For the same or similar object o and for the same or similar effect e, there will be a number of different actions $a_1, a_2, \cdots$:

$$o \begin{array}{c} \xrightarrow{a_1} \\ \xrightarrow{a_2} \\ \vdots \end{array} e.$$

Since these actions $a_1, a_2, \cdots$ have the same or similar function $o \rightarrow e$, namely having the effect e on the object o, we call them *functionally equivalent actions*. Functionally equivalent actions have the same or similar *value* or *meaning* for the agent.

This functional equivalence is learnable by observing others perform various actions and by performing various actions by oneself. Grouping actions that share a function $o \rightarrow e$, the learner obtains a functionally equivalent set A of actions $\{a_i\}$ for which one has experienced (i.e. observed or performed) $o \overset{a_i}{\rightarrow} e$:

$$o \overset{A}{\boxed{\begin{array}{c} \xrightarrow{a_1} \\ \xrightarrow{a_2} \\ \vdots \end{array}}} e.$$

Note that an action $a_i \in A$ may also belong to another group A'; each member a_i of a group A has a weight w_i that represents how relevant a_i is to the function $o \rightarrow e$.

Now consider the following two actions a_{exe} and a_{obs}; a_{exe} is a motor program that I execute *myself*, while a_{obs} is an observation of *his/her* movement. Assume that both a_{exe} and a_{obs} are applied to the same or similar object o and produce the same or similar effect e:

$$o \underset{a_{\text{exe}}}{\overset{a_{\text{obs}}}{\rightrightarrows}} e.$$

Since both have the same or similar function $o \rightarrow e$, the executed action a_{exe} and the observed action a_{obs} are *functionally equivalent*, although they are not directly comparable — one is in the motor space, and the other in the vision space.

3. Learn to mirror

Arbib (2002) proposed a model of learning in the mirror system based on the neurological findings and on his model of language evolution. According to the learning model and his recent model (Oztop and Arbib to appear) the mirror system develops for grasping in the following way:

1. explore a set of basic grasps (i.e. actions: $o \overset{a}{\rightarrow} e$) through trial and error by *myself*;
2. associate visual observations of *my* own hand movement with *my* basic grasps;
3. match visual observations of *his/her* hand movement to *my* basic grasps by using the association above.

The point of their learning model is that the mirror system (a part of F5 in macaques' cerebral cortex) learns the correlation between views of *my* movement and *my* motor programs. The equivalence between *my* hand and *his/her* hand is visually recognized (by STS and 7a/b) in terms of hand posture/movement and spatial relation between hands and objects.

On the other hand, our model described in the previous section learns to correlate *his/her* action with *my* action in the following way:

1. explore a set of *my* actions ($o \overset{a_{\text{exe}}}{\rightarrow} e$) through trial and error by *myself*;
2. explore a set of *his/her* actions ($o \overset{a_{\text{obs}}}{\rightarrow} e$) by observing *his/her* behavior;
3. *my* actions and *his/her* actions are incrementally grouped into functionally equivalent sets, each of which is for a function $o \rightarrow e$.

Our model does not require any visual analysis of hands in order to learn the functional equivalence of actions. Imagine that a blind infant learns the functional equivalence by feeling objects and effects on them; also imagine that a physically challenging infant learns the functional equivalence between the others' hand grasp and the infant's foot grasp.

4. Imitation of novel actions

The mirror system works as a bidirectional mapping between functionally equivalent actions of the self (executable) and others (observable):

$$a_{\text{obs}} \Longleftarrow (o \rightarrow e) \Longrightarrow a_{\text{exe}},$$

which now no longer requires the mediate function $o \rightarrow e$ as a clue for recalling a_{exe} and a_{obs}:

$$a_{\text{obj}} \Longleftrightarrow a_{\text{exe}}.$$

For already known actions one can thus mimic others' body movement.

Imitation is, however, the learning strategy that enables us to acquire *novel* behaviors at the very first sight. The mirror system described above needs a number of executions and observations to learn the mapping between functionally equivalent actions of the self and others. How could instant imitation of novel actions be possible?

We hypothesize that a novel action α can be imitated if α is decomposed into a temporal sequence of already known action units:

$$\alpha = \langle a_1, a_2, \cdots, a_n \rangle,$$

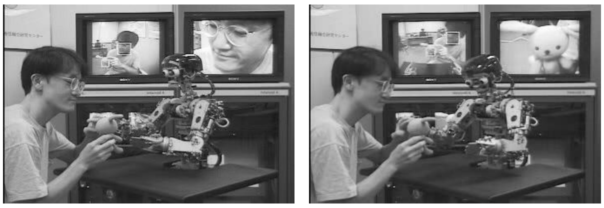

Fig. 2 *Infanoid* engaging in joint attention.

where a_i is an action unit for which a functionally equivalent pair of an observable $a_{i,\mathrm{obs}}$ and an executable $a_{i,\mathrm{exe}}$ is already known:

$$a_{i,\mathrm{obs}} \Longleftrightarrow a_{i,\mathrm{exe}}.$$

Thus the novel action α can be imitated as a temporal sequence[1] of already known action units:

$$\langle a_{1,\mathrm{obs}}, \cdots, a_{n,\mathrm{obs}} \rangle \Longleftrightarrow \langle a_{1,\mathrm{exe}}, \cdots, a_{n,\mathrm{exe}} \rangle.$$

5. Further research

We are currently implementing a robotic system that is able to construct a mirror system through the interaction with physical environment (objects to play with) and social environment (human caregivers to play with). For this purpose we are building a series of humanoid robots *Infanoid*. *Infanoid* possesses approximately the same kinematic structure of the upper half of the body of a three-year-old human infant. The latest version (shown in Fig. 1) has 31 DOFs arranged in a 480-mm-tall upper body.

Infanoid is currently capable of (1) engaging in joint attention with human caregivers, where *Infanoid* reads the direction of the human face and pays attention to the object of interest (as shown in Fig. 2, with the previous version of *Infanoid*), (2) pointing to or reaching out for an object of shared interest, (3) mimicking human vocalization with the lips moving during the joint engagement, as well as spontaneous babbling-like vocalization.

What we need in “learning to mirror” by *Infanoid* is a system for computing affordances of an object of interest. We are planning to first build a heuristic-based affordance responder in order to evaluate our grand scheme and then design an autonomous affordance explorer that best utilizes objects’ affordances through trial and error.

6. Conclusion

We have outlined here our preliminary model of the emergence of a mirror system for imitative learning. The mirror system is *constructed* by correlating *my* own action on an object for producing an effect (i.e. an executable motor program) with *his/her* action on the same or similar object for producing a similar effect (i.e. mainly visual observation). Because both *I* and *he/she* have functionally similar bodies and therefore respond to the object’s affordance in a similar way and evaluate the produced effect in a similar way, the mirror system can learn the relation between executable and observable actions in terms of *functional equivalence*.

Mirroring one’s own action and others’ action enables us to understand empathetically others’ internal states (e.g. intentions). This would play a key role not only in predicting and controlling others’ behavior, which is indispensable for cooperation and competition in a group of individuals, but also in imitative learning, which enables us to share and maintain individual inventions in a group of individuals over generations.

References

Arbib, M. (2002) The mirror system, imitation, and the evolution of language, in Nehaniv, C. and Dautenhahn, K. (eds), *Imitation in Animals and Artifacts*, The MIT Press.

Kozima, H. (2002) Infanoid: a babybot that explores the social environment, in Dautenhahn, K., Bond, A. H., Cañamero, L., and Edmonds, B. (eds), *Socially Intelligent Agents: Creating Relationships with Computers and Robots*, pp. 157–164, Kluwer Academic Publishers.

Oztop, E. and Arbib, M. (to appear) Schema design and implementation of the grasp-related mirror neuron system, *Biological Cybernetics*.

Rizzolatti, G. and Arbib, M. A. (1998) Language within our grasp, *Trends in Neuroscience*, Vol. 21, pp. 188–194.

Tomasello, M. (1999) *The Cultural Origins of Human Cognition*, Harvard University Press.

[1] In the same way, we can also consider imitation of spatial (and spatio-temporal) combination of action units.