What is an Appropriate Theory of Imitation for a Robot Learner?

Joe Saunders, Christopher L. Nehaniv and Kerstin Dautenhahn
Adapative Systems Research Group
University of Hertfordshire
College Lane, Hatfield, UK, AL10 9AB
Email: j.i.saunders,C.L.Nehaniv,K.Dautenhahn@herts.ac.uk

Abstract—When faced with learning new skills or trying to solve problems humans sometimes have a choice, either learn the skill oneself (individual learning) or alternatively learn the skill from someone who already has that knowledge (social learning). Learning from another has the advantage of bypassing a long and costly search process and can therefore speed up learning by exploiting the knowledge of the other. It may also be the case that new and original solutions to problems may never be found when searching for them individually whereas readily available solutions may be found by exploiting the knowledge of another. Imitation is one mechanism which allows this knowledge transfer to take place.

In this paper comparisons of selected theories underlying imitation and how these may be useful from a robotics viewpoint and especially a robot learner are presented. Many of these theories underlie existing robotic implementations and thus to understand the efficacy of these robot learners an understanding of the underlying concepts may prove useful and necessary.

I. INTRODUCTION

Imitative learning, where an agent typically learns new skills by observing and repeating the behaviour of others, is a powerful mechanism for knowledge acquisition and seems to be a key characteristic of human forms of intelligence. Imitation is thought to play a role in the formation of social relationships [1], cultural transmission [2], [3] and language acquisition [4] and seems to be central in the process by which humans become social beings [5]. This process of socialisation may have been the evolutionary reinforcing driver which both led to the high level of imitative skills in humans and to further degrees of socialisation which in turn led to higher imitative abilities [6]. One of the key features of highly developed social skill is the identification of others as people, and specifically as “like me” [7], as well as the ability to be empathic by being able to judge the feelings or internal states of others.

From a robotics and computer science viewpoint, imitation can be considered as an efficient way of learning that might avoid the need for complex programming and reduce the need for robotics experts by providing a straightforward way in which non-specialists can instruct artificial systems to carry out useful tasks. Imitation and its role in the development of social beings is also becoming very important in robotics research as it becomes increasingly apparent that to co-exist with humans robots will require social skills and the ability to empathise with others [8]. Imitation is still however a deep and hard problem and research continues in studying what the neuro-biological features are which support imitation and how imitation can be implemented within artificial systems.

In this paper we focus on this latter point and present some comparisons of three important theories of imitation and how these may considered from a robotics viewpoint. We hope that this will prove useful for those involved in the development of robots by highlighting the main issues that should be considered in an imitative architecture.

II. WHAT’S IMPORTANT FOR ROBOTIC ImitATION?

In this section some of the main proposals emanating from developmental and neuro-psychology will be presented in order to understand various theoretical ideas that are current in imitation research. A comparison of these theories and their power to explain key issues in imitation will be discussed in order to provide a firm basis for the decisions on the robotic mechanisms for social learning. The comparison is based on a set of questions against which the various theories can be compared. Although these questions arise from human specific questions in imitation research they can also be considered in terms of whether the theory can be realistically realised in a robotic implementation.

The theoretic proposals tend to be either specialist in that a specific mechanism for imitation is purported to exist, or generalist, in that existing mechanisms for learning and action control are employed [9]. In this study the following questions are considered to be those which the theoretic proposals should consider:

• How is the correspondence problem addressed or solved?
• How does the theory address the problem of perceptual-motor translation?
• Is the theory compatible with research on mirror neurons?
• Does the theory employ mechanisms compatible with kinesthetic-visual matching, mirror imitation and self-assisted-imitation?
• Can the theory be realistically implemented in a robotic system?

It is not suggested that these are the only issues that a theory of imitation employed in a robotic system should address, however this set of questions considers issues which are at the heart of imitation research and some of which may be key to the design of an artificial mechanism and as such their
interpretation should provide clues as to the effectiveness of the mechanisms proposed.

III. ISSUES IN IMITATION

One debate on human imitation centres on whether imitation is an innate and specialist activity that is present in humans at birth or alternatively whether imitation is a feature which exists as part of a generalist sensorimotor mechanism [10]. The latter explanation being important as it then proposes that imitative behaviour can be learned - thus humans learn how to imitate as well as learn from imitation. From a robotics perspective this is an important distinction as it emphasises the need for an adaptive mechanism providing a facility whereby imitative abilities can be acquired as well as the ability to acquire further task knowledge from existing imitative abilities. This section examines aspects of human imitative behaviour which give rise to this debate.

A. Imitation in Infants and Neonates

It is believed that very young babies exhibit imitative behaviour by facial imitation. Typically this involves tongue protrusion, lip protrusion and mouth opening. These have been studied by a number of researchers (e.g. [11], [12]) and these effects are taken by Meltzoff and Moore [13] as evidence that neo-natal imitation is an innate mechanism which does not necessarily require learning, i.e. humans do not need to learn how to imitate. A review study by Anisfeld [14] however suggested that only tongue protrusion showed any correlation with observation of the model movements. This finding has been taken by Heyes [10] as an indication that tongue protrusion may be an innate releasing mechanism (IRM). Heyes states “If tongue protrusion is the only body movement that newborns can imitate, it is plausible that it is an IRM; an inborn stimulus-response link, wherein the response coincidentally resembles the stimulus from a third party perspective” [10, p. 253]. This implies that imitation may be a generalist activity and therefore require that imitation be learned. However, studies by Meltzoff [7] indicate that over time neonates refine their tongue protrusion skills matching not only protrusions but movements to the right and left of the mouth. This would conversely suggest that tongue protrusion cannot be an IRM. The scientific debate on neonatal imitation is still ongoing.

B. Kinesthetic-visual Matching, Mirror imitation and Self-imitation

Kinesthesia (including somathesis - the feeling of the outline of the body) is the perceptual system that “informs about the position and movement of parts of the body” [15, p. 449]. To experience kinesthesia requires that the body moves. Kinesthetic-visual matching occurs between the kinesthetic feelings and the visual feedback from observing the movement. This movement can be from observing one’s own body movements, observing movements in a mirror or the observation of the movements of another. It can also be considered that the process of self or assisted imitation [16], [17] in animals and/or robots can be another form of sensory-motor matching. This is defined as a multi-modal association of the movement of an individual’s body parts combined with the matching of visual and other environmental stimuli.

C. The Correspondence Problem

At a general level imitation may be hypothesised to involve some form of copying and some mechanism for assessing similarity between the actions of the model and the actions of the imitator. This presents some difficult challenges. How does the imitator match the model’s body with the equivalent parts of its own body especially so if the bodies are different? If this matching is successful how does it match the orientation of actions made by the model? E.g. is a mirroring of actions correct or an exact match? Assuming that this problem is solved, does the imitator match the goals of the model (e.g. start the car) or the actions of the model (e.g. turn the key)? How can the similarity of the mapping be measured? This general problem in imitation is called the “Correspondence Problem” [18] and can be stated in terms of matching states (of the body) and effects (on objects and environment), actions both internal and external to the imitator that transform the state, subgoals and goals that describe configurations of state and granularity as a measure the fineness of the matching attempt. Solving correspondence problems is one of the key challenges in imitation research.

D. Mirror Neurons

Mirror neurons are brain cells discovered in the premotor area of monkey brains which appear to be active when the monkey observes a goal-directed action made by another monkey (or human) and when it executes that action itself [19]. There is also evidence that similar cells exist in human brains in the prefrontal motor cortex and in Broca’s area which is analogous to the premotor F5 area of the money brain. It has been suggested that this reflects a link between non-verbal communication and language via imitation [4], [20] and as the mechanism which allows empathic abilities to exist in humans via the ability to experience what others may be experiencing [21]. It may be that in some animal species the imitative ability of mirror neurons has resulted from exaptation, and thus used by the animal for something other than its original purpose. This being the case, it can then be argued that mirror neurons are not specifically for imitation, but “acquire their properties in the course of ontogeny as a side-effect of the operation of general associative learning and action control processes” [9, p. 489]. This idea is also supported by evidence that mirror neurons for tool use can develop during ontogeny [22] and therefore that learning has a key function in their ontogenesis.

IV. THEORIES OF IMITATION - ASSESSMENT AND EXAMPLES OF ROBOTIC IMPLEMENTATIONS.

In the sections that follow the following selected theories of imitation will be discussed:
• Active Intermodal Mapping (AIM) - (Meltzoff and Moore (1997) [13]).
• Associative Sequence Learning (ASL) - (Heyes and Ray (2000) [23]).
• Extended Ideomotor Theory (EIT) - (Prinz (2005) [24]).

A. Active Intermodal Mapping

One of the puzzles facing researchers in imitation is the ability of the imitator to replicate movements seen in others when they cannot see themselves perform the same action, thus imitating perceptually opaque movements. Meltzoff and Moore [13] address this problem in their theory (see figure 1) by attempting to explain imitation by infants of facial movements in adult models.

![Fig. 1. Active Intermodal Mapping Theory. External and internal perceptions are represented in a common form of organ relations. These are compared. A match implies that the imitation was successful and may indicate a “like me” condition. A mismatch implies that some movement is necessary. The movement is extracted from a directory of coordinated acts previously built from body babbling. (Figure inspired from [13]).](image)

In their proposal a special purpose mechanism is suggested that takes input from observation of the model’s movements and converts these to a ‘supramodal’ encoding as a set of ‘organ relations’. Proprioceptive feedback from the imitator’s own motor outputs are also encoded in this form. The two representations are then compared in a kinesthetic-visual matched ‘goal-directed’ selection process. The equivalence of the relations implying a matching state. Non-equivalence causes a search for a match of the visual perception against prior experiences (in this case body babbling) and the nearest coordinated act is executed. Thus the flow of control forms a feedback loop where the matching experiences inform the imitator of similarity between the model and itself, whereas differences invoke movements in an attempt to gain similarity.

The explanatory power of AIM is best when proposing mechanisms that explain opaque imitation situations, goal-directedness and thus intention. However, it has a number of problems. Firstly, the theory is weak when describing the structure of ‘organ relation’ encoding or how the organ relations are derived from body movements. This makes it difficult to address the correspondence problem issue directly. Secondly, it is based on evidence from neonatal facial imitation such as tongue protrusion, however these movements (as described in section III-A above) may be due to an innate releasing mechanism within the infant and as such inconsistent with imitative learning [10] although Meltzoff and Moore [13] suggest that corrective refinement of movements implies that this must be an in-built imitative function. Thirdly, AIM in suggesting that imitation is a special purpose mechanism implies that a specialised collection of neurons will be at its functional core, which may be consistent with the ideas behind mirror neurons, however it is not consistent with the evidence that imitation is experience dependent in primates and thus a generalist learning/adaptive mechanism must be present, e.g. mirror neurons may support imitation but may not specifically have evolved for imitation. Finally, because there is little detail on how organ relations are encoded the similarity mechanisms underlying mirror imitation and self-imitation are difficult to address.

A common problem with all of the theories of imitation presented in this section is the descriptive nature of the execution of actions. There are many ways of executing an action however the theories are typically vague on the detail involved. A possible strength of AIM from a robotic viewpoint is however that although ‘organ relations’ are weakly defined, the nature of the theory lends itself to implementation in artificial systems. As such, many aspects of AIM have been successfully realised in a number of robotic architectures (see [25]–[29] for examples).

B. Associative Sequence Learning

Associative Sequence Learning (ASL) is based on a set of bi-directional excitatory links between sensor and motor representations of movement units [23].

![Fig. 2. Associative Sequence Learning Theory. Vertical associations are between perceived model actions and the imitators motor programs. Direct vertical associations have no intermediate representation. Indirect links are mediated by another sensory representation (such as a word or phrase). Horizontal associations connect the vertical associations in a chain or sequence of motor actions. (Figure redrawn from [30]).](image)

An action is based on a sequence of these movement units (see figure 2). The horizontal sensory processes (shown at the top of figure 2) are associated. Thus when observing the
movements of the model the visual representations are linked in a chain. It is suggested that this linking occurs either as an activation list where the previous process activates the next or through a context based model where the sensory representations are associated with a time varying control signal [30, p. 518]. Thus the observer learns the sequence of actions made by the model. However in order to replicate these actions ‘vertical’ associations between these sensory inputs and motor representations are necessary, the association formed by Hebbian learning. Motor units are representations of kinesthetic feedback perceived through the execution of the motor action. In ASL, the vertical associations between sensory stimuli and the motor units must be pre-learned. Heyes proposed two types of vertical link: direct links where a movement unit is observed and paired with the sensory input, and indirect links where an alternative sensory stimulus (e.g. a verb or word) is paired with the motor action. Direct links can be formed firstly through self-observation where the imitator observes its own movements, secondly via mirrors where the imitator can observe perceptually opaque movements, thirdly where the imitator observes the model imitating the imitator, fourthly from synchronous action responses of both the model and imitator to a common stimulus and finally from instructor feedback.

ASL is based on kinesthetic-visual matching and explicitly proposes that imitative associations are formed as described above. The mechanism of perceptual-motor translation is implicit in the mechanism described. ASL supports the generalist stance and is therefore consistent with of the idea of mirror neurons being an exaptation. However in terms of addressing the correspondence problem, ASL suggests that the appropriate translations can be addressed as described above (experiencing others, synchronous actions, instructor feedback). However all of the cases present the same logical difficulty (for reviews see [31] and [32]) in that there is nothing which tells the imitator when the actions they see in others correspond to actions in themselves. Thus ASL appears to presuppose that the correspondence problem has already been solved [31]. This leads to one of the practical implementation difficulties posed by ASL in finding ways of resolving the vertical links. Some researchers therefore focus on the horizontal links and then hard-code the vertical associations (see for example [33]). However, Alissandrakis et al. [34] use given metrics to generate different correspondences in the vertical mapping.

C. Extended Ideomotor Theory

Ideomotor theory was originally proposed by William James [35] based on earlier work by Lotze [36] as a way of explaining voluntary behaviour or ‘will’. All actions are held as ‘images’ of the sensory feedback they produce, the sensory stimulation resulting from the internal feedback from the movements of the body as well as external feedback from the effects on the environment. In the Lotze-James approach actions are initiated as a result of imagining their sensory consequences i.e. replicating the sensory feedback experienced.

“Every representation of a movement awakens in some degree the actual movement which is its object; and awakens it in a maximum degree whenever it is not kept from doing so by an antagonistic representation present simultaneously in the mind” [35].

In this way actions are a means of realising intentions (an ideomotor effect), rather than as consequences of sensory stimulation (a sensorimotor effect). Because this approach requires that the effects of the action precede the action itself, Lotze-James proposed that these imagined consequences of action are originally derived from previous learning. Thus the first step is to store a collection of ‘ideas of movement’ from previous learning.

“When a particular movement, having once occurred in a random, reflex or involuntary way, has left an image of itself in the memory, then the movement can be desired again and deliberately willed” [35].

The distinction between voluntary or willed/intentional actions and involuntary actions is not however exact. There are many instances of involuntary actions occurring even though our intention is focused elsewhere. It is thought that this allows skilled actions to be carried out without conscious thought. William James gives an example of this effect where, when dressing for dinner, individuals would end up getting into bed “…merely because that was the habitual issue of the first few movements when performed at a later hour” (quoted in [37]).

![Fig. 3. Extended Ideomotor theory. Ideomotor theory assumes a common coding for the actions and sensory feedback produced. Actions can be invoked from imagining the sensory consequences of that action. Imitation results from a similarity match between perceived actions and and the imagined consequences of that action.](image-url)

The ideomotor effect is thus where an association is made between the action and the consequences of this action. Whenever an action is performed there are a number of perceivable effects, both local such as kinesthetic/affective sensations e.g. the feeling of the movement of one’s fingers and hand, to more distant/environmental effects, such as a car motor starting once the ignition key is turned on with your fingers and hand.
The association theory works in two directions, firstly that with the intention to achieve certain effects, certain actions can be selected. Secondly, that given a certain action, certain effects can be expected. This is similar to the concepts from control engineering using paired inverse and forward models [26, [38], with the inverse model describing the actions required to achieve given goals and the forward models predicting the sensory consequences of given actions (and this could provide a basis for common coding - see below).

Ideomotor theory did not originally address imitation. The extension is however straightforward; if an action can be invoked from thinking of it, then it may also be invoked when perceiving it as performed by someone else. To achieve this the perception of the event must be in some way similar to the effect that has been registered when learning the consequences of an action. One of the key features of ideomotor theory is that it proposes that the original stimulus and the mental images that subsequently come to initiate the voluntary action share a common representational structure and this common coding approach allows the measures of similarity to take place between them.

Extended Ideomotor Theory (EIT) as outlined by Prinz [24] emphasises the role of kinesthetic-visual matching and is also compatible with ideas of mirror imitation and self-imitation. The theory encompasses both the willed/intentional motivations and non-voluntary automatic actions. The issue of perceptual-motor translation is obviated as both share a common representational structure (see [39] for some proposals on how this may be implemented in the brain). Along with ASL, ideomotor theory is a generalist theory and as such is compatible with research on mirror neurons but has little to offer when explaining neonatal imitation. As prior learning/experience is at the heart of ideomotor theory the same mechanisms supporting opaque imitation as proposed in ASL are presumed to exist here (although the mechanism is not outlined by Prinz or others). And, as with ASL, ideomotor theory fails to directly address the correspondence problem. There are (as far as the author knows) few practical implementations of EIT in robotic systems (however see [16] for a partial attempt). Arbib [40] further questions extended ideomotor theory in respect to the emphasis given to effects (and thus intention) and suggests an alternative cyclic approach where neither intention nor pure stimulus-response reaction are given ascendancy. Donald [41] criticises the ideomotor approach based on neuro-physiological grounds in pointing out that the vertebrate motor systems do not follow ideomotor principles in that they are primarily reflexive and thus sensorimotor in nature.

V. SUMMARY AND DISCUSSION

Apart from Active Intermodal Mapping (AIM), the theories above base a large part of their explanatory power on the concepts of prior learning by association and this appears to concur with evidence from developmental studies. Extended Ideomotor Theory (EIT) suggests that learning serves as a mechanism from which the same acts can be (voluntarily) or willed. Associative Sequence Learning (ASL) focuses on how actions can be replicated. The key difference is that EIT emphasises effects whereas ASL tends to focus on actions and states.

The ability of EIT to capture the importance of effects of actions in experience dependent learning and its ability to subsequently use similarity measures to invoke these effects suggests the possibility of imitating actions, states and effects. In the authors’ view this gives this proposal merit in robotic architectures. However, this is not to suggest that AIM or ASL lack merit and in fact aspects of some of these theories have already been demonstrated in various robotic platforms [25–28, [33], it is simply that they are more restricted in this regard.

One major difficulty is however none of the proposals directly address the issues contained in the correspondence problem. It is probably not coincidental that few robotics researchers address this problem either, preferring to ‘hard code’ the mappings instead. Exceptions are firstly the work by Alissandrakis [42] where the correspondence problem is investigated in a simulation of dissimilar embodiments; here however the imitator has access to the full state of the model. And secondly, work by Johnson and Demiris [43] where physical robots are used. In this latter research an implicit perspective transformation is used to place the imitator in the same position as the model. This of course already makes assumptions about relevant mappings between model and imitator and as such avoids part of the problem. However after transformation, the interpretation is based on an abstracting mechanism which operates by relaxing assumptions on motor similarity. All of the predicted states of the imitator’s motor actions (including sub-component actions) are compared with

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AIM</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>ASL</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>EIT</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Fig. 4. Comparison of Theories of Imitation. The table summarises how each theory of imitation deals with the some of the key questions in imitation research.
perceived next state and the most similar model chosen as the imitating motor action. This abstracting mechanism (using forward models) is equivalent to the ideomotor statement of ‘given a certain action, certain effects can be expected’.

This paper has reviewed and compared three important theories of imitation, an important skill for learning and adaptation in natural and artificial systems. We hope that this work will inspire other researchers in developing robots that can learn from each other and human beings by observation and/or interaction.

ACKNOWLEDGMENT

The work described in this paper was partially conducted within the EU Integrated Projects COGNIRON (“The Cognitive Robot Companion”) and Italk I (“Integration and Transfer of Action and Language in Robots”) funded by the European Commission under contract numbers FP6-002020 and P7-214668.

REFERENCES
