Imitation and Social Learning for Synthetic Characters

Daphna Buchsbaum and **Bruce Blumberg** and **Cynthia Breazeal**[1]

Abstract. An increasing amount of evidence suggests that in human infants the ability to learn by watching others, and in particular, the ability to imitate, could be crucial precursors to the development of appropriate social behavior, and ultimately the ability to reason about the thoughts, intents, beliefs, and desires of others [6].

We have created a number of imitative characters and robots [2], the latest of which is Max T. Mouse, an anthropomorphic animated mouse character who is able to observe the actions he sees his friend Morris Mouse performing, and compare them to the actions he knows how to perform himself. This matching process allows Max to accurately imitate Morris's gestures and actions, even when provided with limited synthetic visual input. Furthermore, by using his own perception, motor, and action systems as models for the behavioral and perceptual capabilities of others (a process known as Simulation Theory in the cognitive literature), Max can begin to identify simple goals and motivations for Morris's behavior, an important step towards developing characters with a full theory of mind.

1 INTRODUCTION

Humans (and many other animals), display a remarkably flexible and rich array of social competencies, demonstrating the ability to interpret, predict and react appropriately to the behavior of others, and to engage others in a variety of complex social interactions. We believe that developing systems that have these same sorts of social abilities is a critical step in designing robots, animated characters, and other computer agents, who appear intelligent and capable in their interactions with humans (and each other), and who are intuitive and engaging for humans to interact with.

Since humans provide our inspiration for designing socially intelligent artificial systems, we have approached the challenge by turning to theories of how the ability to interpret the actions and intentions of others, often called theory of mind (ToM), develops in humans. Research in the field of cognitive development suggests that the ability to learn by watching others, and in particular, the ability to imitate, are not only important components of learning new behaviors (or new contexts in which to perform existing behaviors), but could be crucial precursors to the development of appropriate social behavior, and ultimately, theory of mind. In particular, Meltzoff (see [6], [7], [8]) presents a variety of evidence for the presence of imitative abilities in children from very early infancy, and proposes that this capacity could be foundational to more sophisticated social learning, and to ToM. The crux of his hypothesis is that infants' ability to translate the perception of anothers action into the production of their own action provides a basis for learning about self-other similarities, and the connection between behaviors and the mental states producing them.

[1] MIT Media Laboratory, Cambridge MA, USA email: daphna@media.mit.edu

In previous work, we began to explore this hypothesis by implementing a facial imitation architecture for an interactive humanoid robot [2]. In this paper, we present a system that expands upon our prior research, by providing a robust mechanism for observing and imitating whole gestures and movements. Furthermore, the characters presented in this paper are able to use their imitative abilities to bootstrap simple mechanisms for understanding each other's low-level goals and motivations, bringing us a step closer to our goal of creating socially intelligent artificial creatures.

In the next section we briefly explore the cognitive theories motivating our approach in a bit more detail. Subsequently, we describe our imitation architecture, and in particular, look at Max and Morris Mouse, two anthropomorphic animated mouse characters who are able to interact with each other, and observe each other's behavior. We will focus especially on Max's ability to imitate Morris, and on our ongoing research into giving these characters other social learning capabilities, including learning about their environment by observing each other's behavior, and gaining the knowledge necessary to engage in cooperative activities.

2 UNDERSTANDING OTHER'S MINDS

For artificial creatures to possess human-like social intelligence, they must be able to infer the mental states of others (e.g., their thoughts, intents, beliefs, desires, etc.) from observable behavior (e.g., their gestures, facial expressions, speech, actions, etc.). In humans, this competence is referred to as a theory of mind (ToM) [10], folk psychology [5], mindreading [12], or social commonsense [9].

In humans, this ability is accomplished in part by each participant treating the other as a conspecific—viewing the other as being like me. Perceiving similarities between self and other is an important part of the ability to take the role or perspective of another, allowing people to relate to and to empathize with their social partners. This sort of perspective shift may help us to predict and explain others emotions, behaviors and other mental states, and to formulate appropriate responses based on this understanding. For instance, it enables us to infer the intent or goal enacted by anothers behavior—an important skill for enabling richly cooperative behavior.

2.1 Simulation Theory

Simulation Theory (ST) is one of the dominant hypotheses about the nature of the cognitive mechanisms that underlie theory of mind [5], [4]. It can perhaps best be summarized by the clich to know a man is to walk a mile in his shoes. Simulation Theory posits that by simulating another persons actions and the stimuli they are experiencing using our own behavioral and stimulus processing mechanisms, humans can make predictions about the behaviors and mental states of others based on the mental states and behaviors that we would possess in their situation. In short, by thinking as if we were the other person, we can use our own cognitive, behavioral, and motivational systems to understand what is going on in the heads of others.

From a design perspective, Simulation Theory is appealing because it suggests that instead of requiring a separate set of mechanisms for simulating other persons, we can make predictions about others by using our own cognitive mechanisms to recreate how we would think, feel, and act in their situation—thereby providing us some insight into their emotions, beliefs, desires, and intentions, etc. We argue that an ST-based mechanism could also be used by robots and animated characters to understand humans and each other in a similar way. Importantly, it is a strategy that naturally lends itself to representing the internal state of others and of the character itself in comparable terms. This would facilitate an artificial creature's ability to compare its own internal state to that of a person or character it is interacting with, in order to infer their mental states or to learn from observing their behavior. Such theories could provide a foothold for ultimately endowing machines with human-style social skills, learning abilities, and social understanding.

In the following section, we discuss our Simulation Theory-based imitation and movement recognition architecture, which we have developed using two 3D computer animated characters, Max and Morris Mouse.

3 MAX AND MORRIS

Max and Morris are the latest in long line of interactive animated characters developed by the Synthetic Characters Group at the MIT Media Lab [11], [1], [3]. They were built using the Synthetic Characters C5m toolkit, a specialized set of libraries for building autonomous, adaptive characters and robots. The toolkit contains a complete cognitive architecture for synthetic characters, including perception, action, belief, motor and navigation systems, as well as a new, high performance graphics layer for doing Java-based OpenGL 3D Graphics. A brief introduction to a few of these systems will be given here, but it is beyond the scope of this paper to discuss them all in detail (for more information please see [1], [3]).

3.1 The Motor System

For most character architectures, including the one implicit in this work, a creature consists broadly of two components: a behavior system and a motor system. Where the behavior system is responsible for working out what the creature ought to be doing, the motor system is responsible for carrying out the behavior systems requests. The primary task of the motor system for a conventional 3D virtual character is therefore to generate a coordinated series of animations that take the character from where his body is now to where the behavior system would like it to be.

To approach this problem, we have created multi-resolution, directed, weighted graphs, known as *posegraphs*. To create a character's posegraph, source animation material is broken up into *poses* corresponding to key-frames from the animation, and into collections of connected poses known as *movements*. Animations can be generated and played out on the character in real-time by interpolating down a path of connected pose nodes, with edges between nodes representing allowable transitions between poses. The graph represents the possible motion space of a character, and any motor action the character executes can be represented as a path through its posegraph.

Within the posegraph representation, *movements* are of particular importance to us here. Movements generally correspond to things we

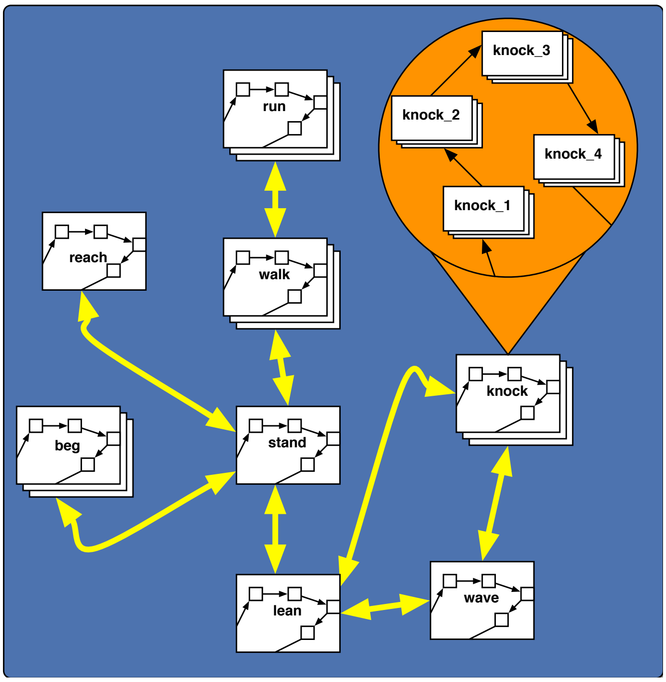

Figure 1. An example graph of movement nodes. Large rectangles represent movements, small squares represent poses. Stacks represent movements and poses created by blending multiple source animations together

might intuitively think of as complete actions (e.g, sitting, jumping, waving), and therefore often match up closely with requests from the behavior system. While the pose representation provides us with greater motor knowledge and flexibility, the movement representation is often a more natural unit to work with. More critically, because movements correspond closely to motor primitives, or to simple behaviors, they also represent the level at which we would like to parse observed actions, in order to identify and imitate them. Therefore, inspired by Simulation Theory, our characters recognize and imitate actions they observe by comparing them with the movements they are capable of performing themselves, a process we will discuss in greater detail in the following section.

3.2 Imitation and Movement Recognition

Max the Mouse is able to observe and imitate his friend Morris's movements, by comparing them to the movements he knows how to perform himself. Max watches Morris through a color-coded synthetic vision system, which uses a graphical camera mounted in Max's head to render the world from Max's perspective. The color-coding allows Max to visually locate and recognize a number of key end-effectors on Morris's body, such as his hands, nose and feet. Currently, Max is hard-wired to know the correspondence between his own effectors and Morris's (e.g. that his right hand is like Morris's right hand), but previous projects have featured characters using learned correspondences [2], and a similar extension is planned for this research.

As Max watches Morris, he roughly parses Morris's visible behavior into individual movements and gestures. Max locates places where Morris was momentarily still, or where he passed through a transitionary pose, such as standing, both of which could signal the beginning or end of an action. Max then tries to identify the observed movement, by comparing it to all the movement representations contained within his own movement graph. To do this, Max compares the trajectories of Morris's effectors to the trajectories his own limbs would take while performing a given movement. This process allows Max to come up with the closest matching motion in his repertoire, using as few as seven visible effectors (as of writing, we have not tested the system using fewer than seven). By performing his best matching movement or gesture, Max can imitate Morris.

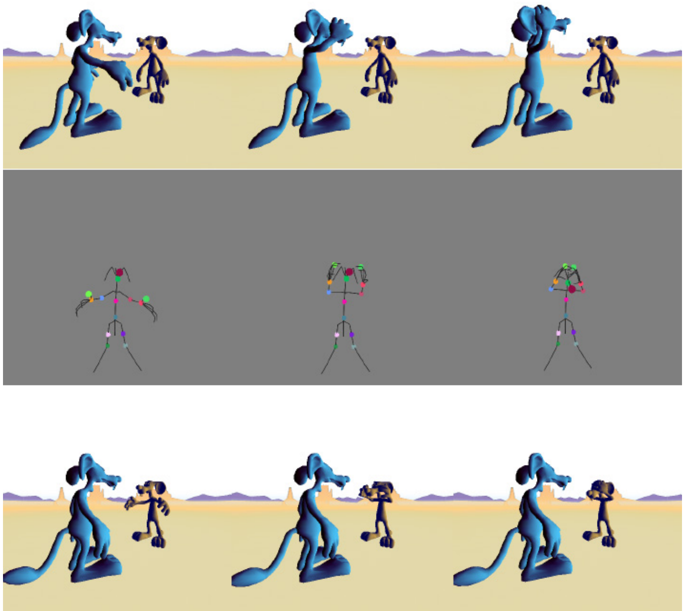

Figure 2. First row: Morris (blue) demonstrates an action (covering his eyes) while Max (brown) watches. Second row: Morris through Maxs eyes. The colored spheres represent key effectors. Third row: Max reproduces Morriss action, by performing the movements in his own repertoire that are closest to what he observed.

3.2.1 Matching Observed Gestures to Movements in the Graph

As Max watches Morris demonstrate a gesture, he represents each frame of observed motion by noting the world-space positions of Morris's effectors relative to Morris's 'root-node' (the center of Morris's body). He then searches his posegraph for the poses (frames) closest to the beginning of the observed action (e.g. poses with similar hand, nose, and foot positions to those hes seen), using the Cartesian distance between corresponding effectors as his distance metric. Max uses these best-matching poses as starting places for searching his posegraph, exploring outward along the edges from these nodes, and discarding paths whose distance from the demonstrated gesture has become too high. Max can then look at the generated path through his graph and see whether it corresponds closely to any of his existing movements, or whether it represents a novel gesture.

One important benefit of using the posegraph to classify observed motion is that it simplifies the problem of dealing with partially observed (or poorly parsed) input. If Max watches Morris jump, but doesnt see the first part of the motion, he will still be able to classify the movement as jumping because the majority of the matching path in his posegraph will be contained within his own jump movement. Conversely, if Max has observed a bit of what Morris was doing before and after jumping, as well as the jump itself, he can use the fact that the entire jump movement was contained within the matching path in his graph to infer that this is the important portion of the observed motion. In general, this graph-based matching process allows observed behaviors to be classified amongst a characters own actions in real-time without needing any previous examples.

4 IDENTIFYING ACTIONS, MOTIVATIONS AND GOALS

Max and Morris both choose their actions using a hierarchically organized action system, composed of individual action units known as action tuples (detailed in [1]). Each action tuple contains an action to perform, trigger contexts in which to perform the action, an optional object to perform the action on, and do-until contexts indicating when the action has been completed. Within the each level of the action hierarchy, tuples compete probabilistically for expression, based on their action and trigger values.

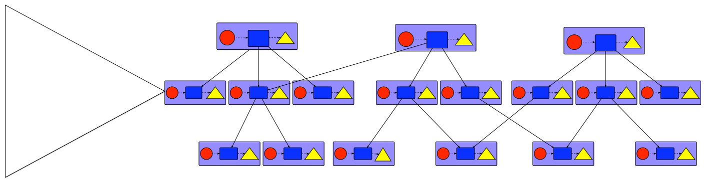

Figure 3. An example action system. Purple rectangles represent tuples. Red circles are trigger contexts, yellow triangles are objects, and blue rectangles are actions (do-until contexts not shown)

4.1 Action Identification

By matching observed gestures and movements to his own, Max is able to imitate Morris. Max can also use this same ability to try and identify which actions he believes Morris is currently performing. Max keeps a record of movement-action correspondences, that is, which action he is generally trying to carry out when he performs a particular movement (e.g. the 'reaching' gesture is most often performed during the 'getting' action). When he sees Morris perform a given movement, he identifies the action tuples it is most likely to be a part of. He then evaluates a subset of the trigger contexts, known as *can-I* triggers, to determine which of these actions was possible under the current circumstances. In this way, Max uses his own action selection and movement generation mechanisms to identify the action that Morris is currently performing.

4.2 Motivations and Goals

Another subset of trigger contexts, known as *should-I* triggers, can be viewed as simple motivationsfor example, a should-I trigger for Maxs eating action is hunger. Similarly, some do-until contexts, known as *success* contexts, can represent low-level goalsMaxs success context for reaching for an object is holding the object in his hands. By searching his own action system for the action that Morris is most likely to be performing, Max can identify likely *should-I* triggers and *success* do-untils for Morriss current actions. For example, if Max sees Morris eat, he can match this with his own eating action, which is triggered by hunger, and know that Morris is probably hungry. Similarly, Max can see Morris reaching for, or jumping to get, an object, and know that Morris's goal is to hold the object in his hands, since that is the success context for Max's own 'get' action. Notice that in this second case, Max does not need to discern the purpose of jumping and reaching separately, since these are both subactions 'get' in his own hierarchy.

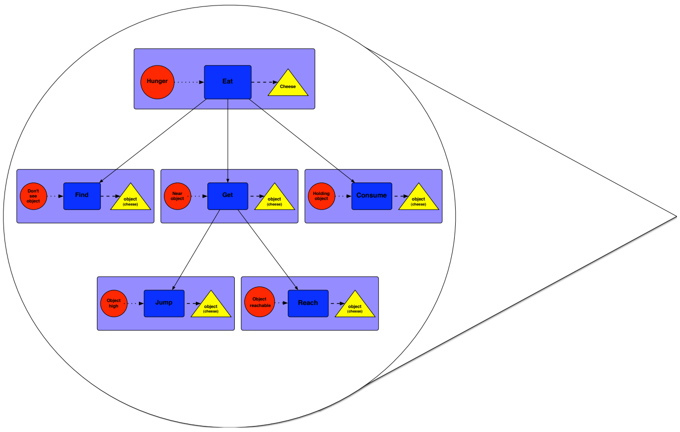

Figure 4. A close up of a motivational subsystem in the action system hierarchy (in this case the hunger subsystem)

We are currently developing mechanisms that allow Max to use the trigger and do-until information from his best matching action in order to interact with Morris in a more socially intelligent wayfor instance, Max might see Morris reaching and help him get the object he is reaching for, bringing him closer to more advanced social behavior such working on cooperative tasks.

5 CONCLUSION

We want to build animated characters and robots capable of rich social interactions with humans and each other, and who are able to learn by observing those around them. This paper presents an approach to creating imitative, interactive characters, inspired by the literature on infant development and by the Simulation Theory view of social cognition. Additionally, it introduces our ongoing work towards creating robots and animated characters who are able to understand simple motivations, goals and intentions, a critical step in creating artificial creatures who are able to interact with humans and each other as socially capable partners.

ACKNOWLEDGEMENTS

We would like to thank the members of the Synthetic Characters and Robotic Life Groups of the MIT Media Lab for their help with, and contributions to, this project.

REFERENCES

[1] B. Blumberg, M. Downie, Y. Ivanov, M. Berlin, M.P. Johnson, and B. Tomlinson, 'Integrated learning for synthetic characters', *ACM Transactions on Computer Graphics*, **21**, 417–426, (2002).

[2] C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby, and B. Blumberg, 'Learning from and about others: Towards using imitation to bootstrap social understanding of robots', *Artificial Life*, **forthcoming**, (2004).

[3] R. Burke, D. Isla, M. Downie, Y. Ivanov, and B. Blumberg. Creature smarts: The art and architecture of a virtual brain. Proceedings of the 2001 Computer Game Developers Conference, 2001.

[4] M. Davies and T. Stone, *Mental Simulation*, Blackwell Publishers, Oxford, 1995.

[5] R. Gordon, 'Folk psychology as simulation', *Mind and Language*, **1**, 158–171, (1993).

[6] A. Meltzoff, *The Human Infant as Imitative Generalist: A 20-Year progress report on infant imitation with implications for comparative psychology*, 347–370, Social Learnign in Animals: The Roots of Culture, Academic Press, New York, 1996.

[7] A. Meltzoff and J. Decety, 'What imitation tells us about social cognition: a rapproachment between developmental psychology and cognitive neuroscience', *Transactions of the Royal Society of London B*, **358**, 491–500, (2003).

[8] A. Meltzoff and A. Gopnik, 'The role of imitation in understanding persons and developing a theory of mind', *Developmental Psychology*, **24**, 470–476, (1993).

[9] M. Meltzoff and M.K. Moore, 'Explaining facial imitation: A theoretical model', *Early Development and Parenting*, **6**, 179–192, (1997).

[10] D. Premack and G. Woodruff, 'Does the chimpanzee have a theory of mind?', *Behavioral and Brain Sciences*, **1**, 515–526, (1978).

[11] B. Tomlinson, M. Downie, M. Berlin, J. Gray, D. Lyons, J. Cochran, and B. Blumberg. Leashing the alphawolves: Mixing user dfirection with autonomous emotion in a pack of semi-autonomous virtual characters. proceedings of the Symposium on Computer Animation, 2002.

[12] A. Whiten and W. Byrne, *Machiavellian Intelligence II: Extensions and Evaluations*, Cambridge University Press, 1997.