Transparency and Socially Guided Machine Learning

Andrea L. Thomaz, Cynthia Breazeal
MIT Media Laboratory
20 Ames St. E15-468
alockerd@media.mit.edu

***Abstract*— In this paper we advocate a paradigm of socially guided machine learning, designing agents that take better advantage of the situated aspects of learning. We augmented a standard Reinforcement Learning agent with the social mechanisms of attention direction and gaze. Experiments with an interactive computer game, deployed over the World Wide Web to over 75 players, show the positive impact of these social aspects. Allowing the human to direct the agent's attention creates a more efficient exploration strategy. Additionally, gaze behavior lets the learning agent improve its own learning environment, using transparency to steer the human's instruction.**

I. Introduction

Robots that operate in the human environment will need to be able to learn new skills and tasks 'on the job' from people. It is important to recognize that while the average consumer is not familiar with machine learning techniques, they are intimately familiar with various forms of social learning (e.g., tutelage, imitation). From a Human-Robot Interaction perspective, this raises two important and related questions. First, how do people want to teach robots? Second, how do we design robots that learn effectively from natural human interaction and instruction? We advocate a paradigm of *socially guided machine learning*, exploring the ways in which machine learning can be designed to more fully take advantage of natural human interaction and tutelage.

We draw our inspiration from Situated Learning Theory, a field of study that looks at the social world of children and how it contributes to their development. A key concept is *scaffolding*, where a teacher provides support such that a learner can achieve something they would not be able to accomplish independently [14], [11].

In a situated learning interaction, the teaching and learning processes are intimately coupled. A good instructor maintains an good mental model of the learner's state (e.g., what is understood so far, what remains confusing or unknown) in order to provide appropriate scaffolding to support the learner's current needs. In particular, attention direction is one of the essential mechanisms that contribute to structuring the learning process [22]. Other scaffolding acts include providing feedback, structuring successive experiences, regulating the complexity of information, and otherwise guiding the learner's exploration. In general, this is a complex process where the teacher dynamically adjusts their support based on the learner's demonstrated skill level and success. The learner, in turn, helps the instructor by making their learning process *transparent* to the teacher through communicative acts (such as facial expressions, gestures, gaze, or vocalizations that reveal understanding, confusion, attention), and by demonstrating their current knowledge and mastery of the task [13], [1]. Through this reciprocal and tightly coupled interaction, the learner and instructor cooperate to simplify the task for the other — making each a more effective partner.

This paper investigates the ways in which the social aspects of learning and the dynamics of the teacher-learner interaction can beneficially impact the performance of a machine learning agent. We use a computer game platform as our experimental testbed, *Sophie's Kitchen*, where players interactively train a virtual robot character to perform a task.

We report two interrelated results that show how social mechanisms can be added to a standard learning process to improve both sides of the teaching-learning partnership. Specifically, we report that

- The ability for the teacher to direct the agent's attention has significant positive effects on several learning performance metrics.
- The ability of the agent to use gaze as a transparency behavior results in measurably better human instruction.

These empirical results illustrate that the ability to use and leverage social skills is far more than a good interface technique. It can positively impact the dynamics of the underlying learning mechanisms to show significant improvements in a real-time interactive learning session.

II. Framework

The situated learning process described above stands in dramatic contrast to typical machine learning scenarios that have traditionally ignored "teachability issues" such as how to make the teaching-learning process interactive or intuitive for a human partner. We advocate a new perspective that reframes the machine learning problem as an interaction between the human and the machine. Fig. 1 sketches this distinction. Typically, the machine learning community has focused on Fig. 1(a): a human provides input to the learning mechanism, which performs its task and provides an output model or classification function.

Our social learning perspective models the complete human-machine system, Fig. 1(b). Adding transparency and augmenting the human input to include guidance, introduce key aspects of social learning, highlighting the reciprocal nature of the teaching-learning partnership. We need a principled theory of the content and dynamics of this tightly coupled process in order to design systems that can learn efficiently and effectively from ordinary users. This approach challenges the research community to consider many new questions.

A. How can the input channels available to the human improve the performance of the teaching-learning system?

We can change the input portion of the typical machine learning process in many ways. Traditionally in supervised learning, training happens off-line where many examples are input in batch. However, a situated teaching interaction may provide benefits over this disembodied, out-of-context method. For instance, a situated learner can take advantage of the natural social cues that the human partner will use (e.g., referencing, attention direction) which may significantly reduce the size of the input space for the machine.

It is important to understand the many ways that natural human social cues can frame the input for a standard machine learning process. This paper explicitly examines the effect of allowing the human to guide the attention of a learner and to provide feedback during its exploration process.

B. How can the output channels of the learning agent improve performance of the teaching-learning system?

The output channels of the learner should also take advantage of the situated nature of learning. In a tightly coupled interaction, a "black box" learning process does nothing to improve the quality and relevance of the instruction. However, transparency of the internal state of the machine could greatly improve the learning experience. By revealing what is known and what is unclear, the machine can guide the teaching process. To be most effective, the machine should use cues that will be intuitive for the human partner [5], [2]. For instance, facial expression, eye gaze, and behavior choices are a significant part of this output channel.

Understanding how both expression and behavior can communicate appropriate levels of the internal state of the learning process is an important issue at hand. We explicitly examines a robot's use of gaze as a transparency behavior.

C. How can the temporal dynamics of input and output improve performance of the teaching-learning system?

Finally we must recognize that these input and output channels interact over time. The dynamics of the interaction can change the nature of the input from the human. An incremental, on-line learning system creates a very different experience for the human than a system that must receive a full set of training examples before its performance can be evaluated. The ability to gauge the system's relative level of knowledge in various situations may help the trainer pick "better" examples for the system. Further, a transparent learner may help a non-expert human teacher answer a difficult question: *when is learning finished?*

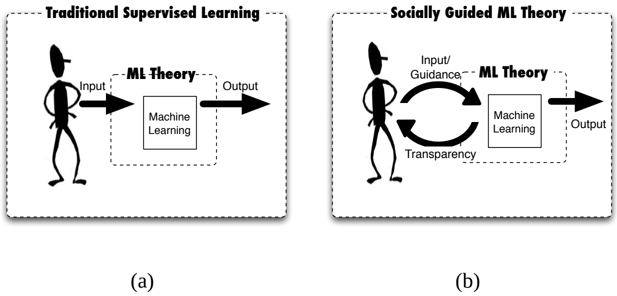


Fig. 1. 1(a) is a standard view of a supervised learning process, analyze input and output a model or classifier, etc. Our approach has the viewpoint of 1(b), including the human teacher, emphasizing that teaching-learning is a two-way process. We add transparency, where the machine learner provides feedback to the human teacher during the learning process; and we augment the human input with guidance. We aim to enhance the performance of the tightly coupled partnership of a machine learner with a human teacher.

This paper examines how the robot's gaze as a transparency behavior interacts with the human's ability to provide guidance during the exploration process, and together how they impact learning performance.

III. Approach

In order to investigate these social aspects of machine learning, we built a computer game platform that allows us to observe people teaching an interactive game character. We deployed the game on the World Wide Web and collected data from over 75 people playing the game. This data allows us to analyze the effects of two important social modifications: the ability of the human to direct the agent's attention, and the ability of the agent to use gaze as a transparency behavior.

A. Platform

Sophie's Kitchen is a Java-based web application that enables the collection of a large amount of human player data. *Sophie's Kitchen* is an object-based State-Action MDP space for a single agent that uses a fixed set of actions on a fixed set of objects. The task scenario is a kitchen world (Fig. 2), where the agent learns to bake a cake; the agent first has to learn to prepare batter for a cake and then to put the batter in the oven. This is a fairly complex task with on the order of 10,000 states and 2 to 7 actions available in each state.

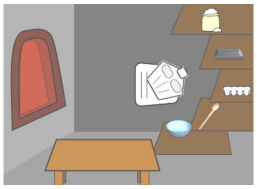
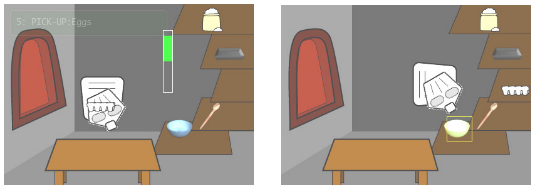


Fig. 2. This is *Sophie's Kitchen*: a brick oven on the left, table in the center, shelf on the right, and the five objects used in the cake baking task.



(a) Feedback message.

(b) Guidance message.

Fig. 3. 3(a), feedback involves left-clicking and dragging the mouse to make a green bar (positive) or red bar (negative). 3(b), guidance involves right-clicking an object of attention, selecting it with the yellow square.

The world has five objects: `Flour`, `Eggs`, a `Spoon`, a `Bowl` (with five object states: `empty`, `flour`, `eggs`, `unstirred`, `stirred`), and a `Tray` (with three object states: `empty`, `batter`, `baked`). The world has four locations: `Shelf`, `Table`, `Oven`, `Agent` (i.e., the agent in the center surrounded by a shelf, table and oven).

Sophie has four base actions with arguments as follows: the locations are arranged in a ring, the agent can `GO left` or `right`; she can `PICK-UP` any object that is in her current location; she can `PUT-DOWN` any object in her possession; and she can `USE` any object in her possession on any object in her current location. Each action advances the world state. For example, executing `PICK-UP <Flour>` changes the state of the world such that the `Flour` is in location `Agent`. The agent can hold only one object at a time. `USE`ing an ingredient on the `Bowl` puts that ingredient in it; using the `Spoon` on the `unstirred Bowl` transitions its state to `stirred`, etc.

In the initial state, all objects are on the `Shelf`, and the agent faces the `Shelf`. A successful task completion includes putting flour and eggs in the bowl, stirring the ingredients using the spoon, transferring the batter into the tray, and putting the tray in the oven.

B. Learning algorithm

Our goal with this experimental platform is to show the differential effects of social mechanisms on a standard, widely understood, machine learning algorithm. Our primary interest is in the on-line scenario of learning via self-generated experience over time. Thus we choose as our standard algorithm the Reinforcement Learning (RL) paradigm. The algorithm we implemented for the experiment presented in this paper is a standard Q-Learning algorithm (learning rate $\alpha = .3$ and discount factor $\gamma = .75$) [21].

The goal state (`tray` in the `oven`, `baked`) has a positive reward ($r = 1$), every other state has an inherent small negative reward ($r = -.04$). Also, some end states are so-called *disaster* states in that there is no way to recover from them (for example—putting the `eggs` in the `oven`). These result in a negative reward ($r = -1$), the termination of the current trial episode, and a transition to the initial state.

C. Feedback Interface

Sophie's Kitchen has an interactive reward interface. Using the mouse, a human trainer can—at any point in the operation of the agent—award a scalar reward signal $r \in [-1, 1]$. This reward is in addition to the environmental rewards just described. The user receives visual feedback enabling them to tune the reward signal before sending it to the agent. Choosing and sending the reward does not halt the progress of the agent, which runs asynchronously to the interactive human reward. Abstractly, the Q-Learning algorithm is continually going through the loop: `select-action`, `take-action`, `sense-reward`, `update-values`. We introduce a delay in the `sense-reward` step to give the human time to provide feedback.

D. Guidance Interface

We want to explore the effects of allowing the teacher to direct the attention of the agent; giving the human the ability to directly influence the action selection and bias the exploration strategy. To accomplish this, we added a separate guidance channel of communication. Clicking the right mouse button draws an outline of a yellow square. When the yellow square is administered on top of an object, this communicates a guidance message to the learning agent where the content of the message is the object. Figure 3(b) shows the player guiding Sophie to pay attention to the bowl.

To incorporate guidance, we modify the Q-Learning algorithm, introducing a pre-action phase. In the pre-action phase, the agent registers guidance communication to bias action selection; in the post-action phase the agent uses the reward channel in the standard way to evaluate that action and update a policy. Thus, the learning process becomes: `sense-guidance`, `select-action`, `take-action`, `sense-reward`, `update-values`.

The agent waits for guidance messages during the `sense-guidance` step (we introduce a short delay to allow the teacher time to administer guidance). A guidance message is in the form `guidance[object]`; the agent saves this `object` as the `guidance-object`. During the `select-action` step, the default behavior (a standard approach) chooses randomly between the set of actions with the highest Q-values, within a bound β. If a guidance message was received, the agent will *instead* choose randomly between the set of actions that use the `guidance-object`.

E. Gazing Behavior

We also explore gaze as a means of making the learning process more transparent to the human teacher. Gaze requires that the learning agent have a physical/graphical embodiment that can be understood by the human as having a forward heading. In general, gaze precedes an action and communicates something about the action that is going to follow.

Recall our guidance Q-Learning loop: `sense-guidance`, `select-action`, `take-action`, `sense-reward`, `update-values`. The gaze behavior executes during the `sense-guidance` phase. The learning agent finds the set of actions, A, with the highest Q-values, within a bound β. $\forall a \in A$, the learning agent gazes for 1 second at the `object-of-attention` of a (if it has one). This gazing behavior during the pre-action phase communicates a level of uncertainty through the amount of gazing that precedes an action. It introduces an additional delay (proportional to uncertainty) prior to `select-action`, both soliciting and providing the opportunity for guidance messages from the human. This also communicates overall task certainty or confidence as the agent will speed up when every set, A, has a single action. We expect this transparency to improve the teacher's model of the learner, creating a more understandable interaction for the human and a better learning environment for the agent.

F. Experimental Design

We deployed the *Sophie's Kitchen* game on the World Wide Web. Participants were asked to play a computer game, in which their goal was to get the virtual robot to learn how to bake a cake on her own. Participants were told they could not tell Sophie what actions to do, nor could they do any actions directly. They were only able to send Sophie various messages with the mouse to help her learn the task. Depending on their test condition, subjects were given instructions on administering feedback and guidance.

Each of the 75 participants, played the game once in one of the following three test conditions:

- `Feedback`: Players used only the feedback communication channel.
- `Guidance`: Players had both the feedback and the guidance channels of communication.
- `Gaze-guide`: Players had the feedback and guidance channels. Additionally, the agent used the gaze behavior.

The system maintained an activity log and recorded time step and real time of each of the following: state transitions, actions, human rewards, guidance messages and objects, gaze actions, disasters, and goals. We then analyzed these game logs to test the following two hypotheses:

- **Hypothesis 1**: Teachers can use attention direction as a form of guidance, to improve a learning interaction.
- **Hypothesis 2**: Learners can help shape their learning environment by communicating aspects of the internal process, particularly that the gaze behavior will improve a teacher's guidance instruction.

IV. Results

A. Guidance Improves Learning

To evaluate the effects of the guidance feature we compare the game logs from players in the `guidance` condition to those in the `feedback` condition with a series of 1-tailed t-tests (summary in Table I).

`Guidance` players were faster than `feedback`. The number of training trials needed to learn the task was 48.8% less, $t(26) = 2.68, p =< .01$; and the number actions needed to learn the task was 54.9% less, $t(26) = 2.91, p < .01$. Thus the ability for the human teacher to guide the agent's attention to appropriate objects at appropriate times creates a significantly more efficient learning interaction.

In the `guidance` condition the number of unique states visited was 49.6% less, $t(26) = 5.64, p < .001$. Thus, attention direction helps the human teacher keep the exploration of the agent in the most useful part of the task space. This is a particularly important result since that the ability to deal with large state spaces has long been a criticism of RL. A human partner may let the algorithm overcome this challenge.

And finally the `guidance` condition provided a more successful training experience. The number of trials ending in failure was 37.5% less, $t(26) = 2.61, p < .01$; and the number of failed trials before the first successful trial was 41.2% less, $t(26) = 2.82, p < .01$. A more successful training experience is particularly desirable when the learning agent is a robot that may not be able to withstand very many failure conditions. Additionally, a successful learning interaction, especially reaching the first successful attempt sooner, may help the human teacher feel that progress is being made and prolong their engagement in the process.

TABLE I

EFFECTS OF GUIDANCE ON THE LEARNING PERFORMANCE. (F = FAILED TRIALS, G = FIRST SUCCESS).

Measure	Mean guidance	Mean no guidance	t(26)	p
# trials	14.6	28.52	2.68	<.01
# actions	368	816.44	2.91	<.01
# states	62.7	124.44	5.64	<.001
# F	11.8	18.89	2.61	<.01
# F before G	11	18.7	2.82	<.01

B. Gaze Improves Guidance

Our second hypothesis was the gaze behavior serves as a transparency device to help the human understand when the agent did (and did not) need their guidance instruction. To evaluate this, we analyzed the game logs from players that had the `guidance` condition versus those that had the `gaze-guide` condition. These results are summarized in Table II. Note that the players that did not have the gaze behavior still had ample opportunity to administer guidance; however, the time that the agent waits is uniform throughout.

We look at the timing of each player's guidance communication and separate their communication into two segments, the percentage of guidance that was given when the number of action choices was $>= 3$ (high uncertainty), and when choices were $<= 3$ (low uncertainty), these are overlapping classes. The percentage of guidance when the agent had low uncertainty decreased for players in the `gaze-guide` condition, $t(51) = -2.22, p = .015$. And conversely the percentage of guidance when the agent had high uncertainty increased from the `guidance` to the `gaze-guide` condition, $t(51) = 1.96, p = .027$. Thus, when the agent uses the gaze behavior to indicate which actions it is considering, the human trainers do a better job matching their instruction to the needs of the agent throughout the training session.

TABLE II

THE EFFECT OF GAZE ON GUIDANCE. COMPARED TO THE GUIDANCE DISTRIBUTION WITHOUT GAZE, THE GAZE CONDITION CAUSED A DECREASE WITH CHOICES $<= 3$, INCREASE WITH CHOICES $>= 3$.

Measure	Mean % gaze	Mean % no gaze	t(51)	p
guidance when # choices <= 3	79	85	-2.22	<.05
guidance when # choices >= 3	48	36	1.96	<.05

V. DISCUSSION

In this research we are promoting a social learning view of machine learning, emphasizing the *interactive* elements in teaching. There are inherently two sides to an interaction, and our approach aims to enhance standard machine learning algorithms from both perspectives of this interaction.

Allowing the human teacher to administer guidance in addition to feedback improves learning performance across a number of dimensions. The agent is able to learn tasks using fewer actions over fewer trials. It has a more efficient exploration strategy that wasted less time in irrelevant states. We argue that a less random and more sensible exploration will lead to more understandable and teachable agents. Guidance also led to fewer failed trials and less time to the first successful trial. This is a particularly important improvement in that it implies a less frustrating teaching experience, which in turn creates a more engaging interaction for the human.

This work offers a concrete example that the *transparency* of the agent's behavior to the human can improve its learning environment. When the learning agent uses gazing behaviors to reveal its uncertainties and potential next actions, people were significantly better at providing more guidance when it was needed and less when it was not. Gaze is just one such transparency device, the exploration of various devices and their relation to the learning process is part of our future work. Additionally these transparency behaviors serve to boost the overall believability of the agent. The issue of believability has been addressed in the animation, video game, and autonomous agent literature for the purpose of creating emotionally engaging characters [20], [3]. One contribution of this work is to show how believability relates to teachable characters to improve the experience of the human and the learning performance of the agent.

Numerous prior works have explored learning agents (virtual or robotic) that can be interactively trained by people. Many of these works are inspired by animal or human learning. For instance, game characters that the human player can shape through interaction have been successfully incorporated into a few computer games [9], [17], [19]. Breazeal *et al.* have demonstrated aspects of collaboration and social learning on a humanoid robot, using social cues to guide instruction [6]. Animal training techniques and human tutelage have been explored in several robotic agents [12], [16], [18], [15]. As a software agent example, Blumberg's virtual dog character can be taught via clicker training, and behavior can be shaped by a human teacher [4].

Many of these prior works agree with our situated learning paradigm for machines, and have emphasized that an artificial agent should use social techniques to create a better interface for a human partner. This work goes beyond gleaning inspiration from natural forms of social learning and teaching to formalize this inspiration and empirically ground it in observed human teaching behavior through extensive user studies. Thus, another contribution of this work is empirical evidence that social guidance and transparency create a good interface for a human partner, *and* can create a better learning environment and significantly benefit learning performance.

Finally, the scenario of human input has received some attention in the machine learning community. There has been work on computational models of teacher/learner pairs [10]. Active learning and algorithms that learn with queries begin to address interactive aspects of a teacher/learner pair [8]. Queries can be viewed as a type of transparency into the learning process, but in these approaches this does not steer subsequent input from a teacher. Instead, through its queries, the algorithm is in control of the interaction. Cohn *et al.* present a semi-supervised clustering algorithm that uses a human teaching interaction, but the balance of control falls to the human (i.e., to iteratively provide feedback and examples to a clustering algorithm which presents revised clusters) [7].

Thus, prior works have addressed how human input can theoretically impact a learning algorithm. In contrast, this work addresses the nature of *real* people as teachers; our ground truth evaluation is the performance of the machine learner with non-expert human teachers. Whereas prior works typically lend control either to the machine or the human, our contribution is the focus on how a machine learner can use transparency behaviors to steer the instruction it receives from a human, creating more reciprocal control of the interaction.

VI. Conclusion

This work shows that designing for the complete human-machine learning system creates a more successful robot learner. We acknowledge that teaching-learning is a reciprocal and dynamic process, and have augmented a standard RL agent with social mechanisms. Our experiments with an interactive computer game character show significant improvement in a real-time interactive learning session with non-expert human teachers. Our results show that allowing the human to guide the learner's exploration by directing its attention creates a more robust and efficient learning strategy. Second, our results show that the learning agent can improve its own learning environment by using transparency behaviors. Revealing information about the possible next actions the agent is considering with its gaze behavior significantly improved the timing of the teacher's guidance messages.

Acknowledgements

This work is funded in part by the Media Lab corporate sponsors of the Things that Think and Digital Life consortia.

References

[1] M. Argyle, R. Ingham, and M. McCallin. The different functions of gaze. *Semiotica*, pages 19–32, 1973.

[2] R. Arkin, M. Fujita, T. Takagi, and R. Hasegawa. An ethological and emotional basis for human-robot interaction. In *Proceedings of the Conference on Robotics and Autonomous Systems*, 2003.

[3] J. Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1997.

[4] B. Blumberg, M. Downie, Y. Ivanov, M. Berlin, M.P. Johnson, and B. Tomlinson. Integrated learning for interactive synthetic characters. In *Proceedings of the ACM SIGGRAPH*, 2002.

[5] C. Breazeal. *Designing Sociable Robots*. MIT Press, Cambridge, MA, 2002.

[6] C. Breazeal, A. Brooks, J. Gray, G. Hoffman, J. Lieberman, H. Lee, A. Lockerd, and D. Mulanda. Tutelage and collaboration for humanoid robots. *International Journal of Humanoid Robotics*, 1(2), 2004.

[7] D. Cohn, R. Caruana, and A. McCallum. Semi-supervised clustering with user feedback, 2003.

[8] D. Cohn, Z. Ghahramani, and M. Jordan. Active learning with statistical models. In G. Tesauro, D. Touretzky, and J. Alspector, editors, *Advances in Neural Information Processing*, volume 7. Morgan Kaufmann, 1995.

[9] R. Evans. Varieties of learning. In S. Rabin, editor, *AI Game Programming Wisdom*, pages 567–578. Charles River Media, Hingham, MA, 2002.

[10] Sally A. Goldman and H. David Mathias. Teaching a smarter learner. *Journal of Computer and System Sciences*, 52(2):255–267, 1996.

[11] P. M. Greenfield. Theory of the teacher in learning activities of everyday life. In B. Rogoff and J. Lave, editors, *Everyday cognition: its development in social context*. Harvard University Press, Cambridge, MA, 1984.

[12] F. Kaplan, P-Y. Oudeyer, E. Kubinyi, and A. Miklosi. Robotic clicker training. *Robotics and Autonomous Systems*, 38(3-4):197–206, 2002.

[13] R. M. Krauss, Y. Chen, and P. Chawla. Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In M. Zanna, editor, *Advances in experimental social psychology*, pages 389–450. Tampa: Academic Press, 1996.

[14] Ed. M. Cole L. S. Vygotsky. *Mind in society: the development of higher psychological processes*. Harvard University Press, Cambridge, MA, 1978.

[15] A. Lockerd and C. Breazeal. Tutelage and socially guided robot learning. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2004.

[16] L. M. Saksida, S. M. Raymond, and D. S. Touretzky. Shaping robot behavior using principles from instrumental conditioning. *Robotics and Autonomous Systems*, 22(3/4):231, 1998.

[17] K. O. Stanley, B. D. Bryant, and R. Miikkulainen. Evolving neural network agents in the nero video game. In *Proceedings of IEEE 2005 Symposium on Computational Intelligence and Games (CIG'05)*, 2005.

[18] L. Steels and F. Kaplan. Aibo's first words: The social learning of language and meaning. *Evolution of Communication*, 4(1):3–32, 2001.

[19] A. Stern, A. Frank, and B. Resner. Virtual petz (video session): a hybrid approach to creating autonomous, lifelike dogz and catz. In *AGENTS '98: Proceedings of the second international conference on Autonomous agents*, pages 334–335, New York, NY, USA, 1998. ACM Press.

[20] F. Thomas and O. Johnson. *Disney Animation: The Illusion of Life*. Abbeville Press, New York, 1981.

[21] C. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8(3):279–292, 1992.

[22] J. V. Wertsch, N. Minick, and F. J. Arns. Creation of context in joint problem solving. In B. Rogoff and J. Lave, editors, *Everyday cognition: its development in social context*. Harvard University Press, Cambridge, MA, 1984.