Robot Learning via Socially Guided Exploration

Andrea L. Thomaz
MIT Media Laboratory
Cambridge, MA 02139
Email: alockerd@media.mit.edu

Cynthia Breazeal
MIT Media Laboratory
Cambridge, MA 02139
Email: cynthiab@media.mit.edu

***Abstract*— We present a learning mechanism, Socially Guided Exploration, in which a robot learns new tasks through a combination of self-exploration and social interaction. The system's motivational drives (novelty, mastery), along with social scaffolding from a human partner, bias behavior to create learning opportunities for a Reinforcement Learning mechanism. The system is able to learn on its own, but can flexibly use the guidance of a human partner to improve performance. An initial experiment shows how a human shapes the learning process through suggesting actions, drawing attention to goal states, and arranging the environment to encourage generalization.**

I. Introduction

Enabling a human to efficiently transfer knowledge and skills to a machine has inspired decades of research. When this work is viewed along a *guidance-exploration* spectrum, an interesting dichotomy appears. Prior works that incorporate human input into a Machine Learning process tend to maintain a constant level of human involvement.

Several are highly dependent on *guidance*, learning nothing without human interaction (e.g., learning by demonstration [1], [22] or by tutelage [9], [12]). In these, the learner does little if any exploration. The teacher must learn how to interact with the machine and know precisely how it needs to perform the task.

Other approaches are almost entirely *exploration* based. For example, letting a human control a reinforcement learner's reward [2], [7], [14], or provide advice [6], [10], or control the agent during training [17]. Exploration approaches have the benefit that learning does not require the human's undivided attention, but often they do not take enough advantage of the human partner and again usually require the human to learn how to interact with the machine.

Our research is similarly motivated by the idea that robotic agents that operate in human environments will need the ability to learn new skills 'on the job' from everyday people (that are not familiar with Machine Learning techniques). However, we observe that a social learner must be able to move flexibly along this *guidance-exploration* spectrum. It should be able to explore and learn on its own, but also take full advantage of a human partner's guidance when available.

In this paper we present a robot learning system—Socially Guided Exploration. The robot is able to frame its own learning problems through a combination of internal motivation and human guidance. Self-motivated exploration creates learning opportunities for a Reinforcement Learning mechanism. The system defines its own goals, learns action policies for those goals, and generalizes this task representation over time. This works within an integrated system of social scaffolding mechanisms and transparency devices that naturally afford human guidance throughout the learning process.

Fig. 1. Leonardo (and simulated Leo) in his workspace with toy boxes.

II. Robot Platform

Our research platfrom is Leonardo ("Leo"), a 65 degree of freedom robot specifically designed for human social interaction (Fig. 1). Leo has speech and vision sensory inputs and uses gestures and facial expressions for social communication. Leo can visually detect objects in the workspace, humans and their head pose [11], and hands pointing to objects. The speech understanding system is based on Sphinx, and has a limited grammar to facilitate accuracy.

The cognitive system extends the `C5M` architecture [2]. The Perception and Belief Systems are particularly relevant to the learning abilities described in this paper.[1] Every time step, the robot has observations from its various sensory processes, $O = \{o_1, .., o_k\}$. The Perception System is a set of *percepts* $P = \{p_1, ..., p_n\}$. Each $p \in P$ is a classification function, such that $p(o) = m$ where $m \in [0, 1]$ is a match value. The Belief System maintains the *belief* set B by integrating these percepts into discrete object representations (based on spatial relationships and various similarity metrics). Figure 2 shows an example of some sensory data leading to five percepts with $m > 0$, that result in two beliefs in B. In this paper, a "state" s refers to a snapshot of the belief set B at a particular time, and S refers to the theoretical set of all possible states. Let $A = \{a_1, ..., a_i\}$ be the set of Leo's basic actions.

[1] For full technical details of the Perception and Belief Systems see [4].

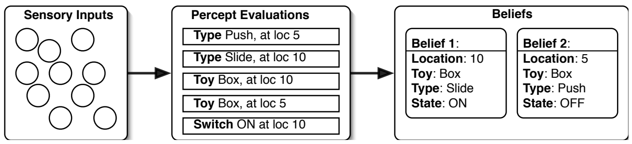

Fig. 2. Sensory input is classified by percepts and then merged into discrete object representations. In this timestep, 5 percepts yield 2 object beliefs.

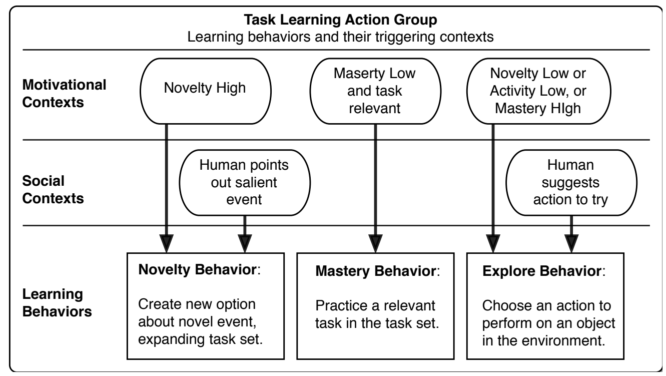

Fig. 3. The three learning behaviors and their social/motivational contexts.

The Socially Guided Exploration system builds on these existing mechanisms, adding capabilities for representing and learning goal-oriented tasks, self-motivated exploratory behavior, and expression/gesture capabilities to support a collaborative dialog with a human teacher.

III. Socially Guided Exploration

In most Machine Learning systems, learning is an explicit activity; the system is designed to learn a particular thing at a particular time. In human learning, on the other hand, learning is a part of all activity; there is a motivation for learning, a drive to know more about the environment, and an ability to seek out the expertise of others. Children explore and learn on their own, but in the presence of a teacher they can take advantage of the social cues and communicative acts provided to accomplish more (also known as *social scaffolding* [8]). A teacher often guides a learner by providing timely feedback, luring them to perform desired behaviors, and controlling the environment so the appropriate cues are salient, thereby making the learning process more effective. This is the primary inspiration for the Socially Guided Exploration system; this section highlights the key implementation details: the Motivation System, learning behaviors, goal-oriented task representation, transparency devices and social scaffolding mechanisms.

A. Motivational Drives for Learning

Living systems work to keep certain critical features within a bounded range through a process of behavioral homeostasis (e.g., food, water, temperature). If a parameter falls out of range, the animal becomes motivated to behave in a way that brings it back into the desired range.

Recently, this concept has inspired work on internal motivations for a Reinforcement Learning (RL) agent [13], [15], [16]. These works use a measure of novelty or certainty as intrinsic reward for a controller. Thus, an action that leads to a prediction error results in rewards that encourage focus on that portion of the space. Our approach is in a similar vein, but rather than contribute to the reward directly, Leo's internal motivations trigger learning behaviors that help the system arbitrate between learning a new task, practicing a learned task, and exploring the environment. Additionally, prior works in "motivated" RL have relied on a single drive (novelty/curiosity). In this work we introduce a mastery drive and demonstrate the benefits of the interplay between novelty and mastery in an agent's learning behavior.

Leo's Motivation System (based on prior work [3]) is designed to guide a learning mechanism. Inspired by natural systems, it has two motivational drives, novelty and mastery. Each drive has a range $[0, 1]$, initial value of 0.5, a tendency to drift to 0.0, and a drift magnitude of 0.001 (max change in a time step). The Motivation System maintains the drive values based on the status of the internal and external environment:

The **Novelty Drive** is an indication of the unfamiliarity of recent events. Every state transition will cause the novelty drive to rise for an amount of time related to the degree of the change, d_{chg}, based on the event's frequency: $d_{chg}(s_1, s_2) = \frac{1}{frequency(s_1,s_2)}$. An event causes the novelty drive to drift towards its maximum value for a period, $t = d_{chg}(s_1, s_2)t_{max}$. The maximum effect time, t_{max}, is 30 seconds.

The **Mastery Drive** reflects the current system confidence of the learned task set. Mastery is the average confidence of the tasks that are relevant in (i.e., can be initiated from) the current state, s. A task's confidence is the the number of successful attempts over the total task attempts made.

B. Learning Behaviors for Motivational & Social Contexts

The Task Learning Action Group is the piece of the Socially Guided Exploration system responsible for identifying and responding to learning opportunities in the environment. It maintains the set of known tasks ($Tasks$), and has three competing learning behaviors that respond to social and motivational learning contexts. Figure 3 is an overview of the behaviors and their internal/external triggering contexts.

1) Novelty behavior: One purpose of the novelty drive is to encourage the system to better understand new events, expanding the $Tasks$ set. Thus, a significant rise in the novelty drive makes the Novelty behavior available for activation. Additionally, this behavior may be activated due to a social context, when the human points out an event (e.g., "Look Leo, it's `TaskName-X`."). Once activated, the Novelty behavior tries to create a new task. It makes a goal representation of the most recent state transition (s_1, a, s_2), and if there is not a $T \in Tasks$ with this goal, then a new task is created. Task creation, expansion, and generalization are covered next in Sections III-C, III-D, & III-E.

2) Mastery behavior: The purpose of the mastery drive is to cause the system to become confident in the environment, fleshing out the representations in the $Tasks$ set. When the mastery drive is low and any tasks are relevant in the current state, the Mastery behavior may be activated. This behavior randomly selects a relevant task, executes it, and updates the confidence based on success in reaching the goal.

3) Explore behavior: Both motivational drives also work to encourage exploration. The Explore behavior becomes available when novelty is low, encouraging the system to seek out the unexpected. Exploration is also triggered when mastery is high; even if a known task is relevant, the system is biased to try to expand the $Tasks$ set once confidence is high. Additionally, social interaction can trigger the Explore behavior, for example if the human suggests an action (e.g., "Leo, try to `Act-X` the `Obj-Y`."). When the Explore behavior is activated, it first tries to do any human-suggested action if possible. Otherwise, the Explore behavior selects from the actions it can do in the current state, with a minimum frequency requirement. Once the action is completed, if it was a human-suggested action, the robot's attention is biased to look to the human. This acknowledges the suggested action and provides an opportunity for feedback.

C. Task and Goal Representation

These three behaviors result in a mechanism that learns object-oriented tasks. Tasks and their goals are represented with *Task Option Policies*. This name reflects its similarity to the Options approach in Reinforcement Learning [20].

Goals encode what must hold true to consider the task achieved. Specifically, a goal $G = \{x_1, ..., x_y\}$ where every $x \in G$ represents a belief that that changed over the task by grouping the belief's percepts into *expectation* percepts (indicating an expected feature value), and *criteria* percepts (indicating which beliefs to apply this expecation to). [2]

Each $T \in Tasks$ is a Task Option Policy, and is defined by a variation of the three Options constructs: I, π, β. To define these we use two subsets of states related to the task. Let $S_{task} \subset S$ be the states in which the task is relevant but not achieved, and $S_{goal} \subset S$ be the states in which the goal is achieved. Then, a Task Option Policy is defined by:

- $\pi' : S_{task} \times A \rightarrow [0, 1]$; estimates a value for (s, a) pairs in relation to achieving the task goal, G.
- $\beta' : S_{goal}$; represents all of the states in which this task terminates because G is true.
- $I' = S_{task}$; represents the initiation set–the task can be initiated in any state for which it has a policy of action.

A task can be executed (is relevant) when the current state is in S_{task}. During execution, actions are chosen according to

[2]This goal construct is also used in prior work, [4], [9].

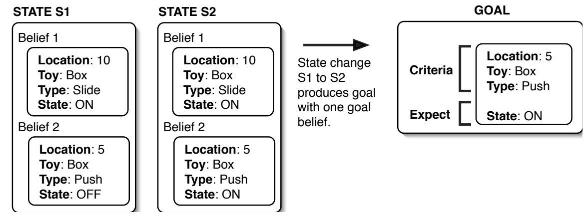

Fig. 4. A simple example of creating a goal from a state change.

π' until the current state is in S_{goal} (with some probability of terminating early). A state s achieves the goal if: $\forall x \in G$, if any belief b in s matches all the $criteria \in x$, then b also matches all the $expectation \in x$.

D. Task Learning

New tasks are created by the Novelty behavior. First, it makes a potential goal state G from the most recent state change, (s_1, a, s_2), with a representation, x, for each belief in s_1 that changed in $s_1 \rightarrow s_2$. Any percept that changed is an expectation, the rest are criteria (e.g., see Fig. 4).

If there does not exist a $T \in Tasks$ with goal G, then a new Task Option Policy, T_{new}, is created. The S_{task} of T_{new} is initialized with the initiation state s_1, and π' is initialized with default values $q = .1$ for all actions from s_1. Then, the system takes into account the experience of (s_1, a, s_2), and (s_1, a) gets a higher value since s_2 is the goal.

Each $T \in Tasks$ can learn and expand from every experience (also referred to as intra-option learning [19]). Every action is an experience, (s_1, a, s_2); and each $T \in Tasks$ has the opportunity to extend its set S_{task} and update its π' based on this experience. To update π', rather than rely solely on external rewards from the environment, the system estimates the reward function based on the task's goal: $r = 1$ if the goal is true in s_2, otherwise $r = 0$.

E. Task Generalization

In addition to expanding initiation sets and updating value estimates for tasks, the system tries to generalize tasks over time. It works to generalize both the state representations in S_{task} and the goal representation G for all $T \in Tasks$.

Given two different tasks T_1 and T_2, the generalization mechanism attempts to combine them into a more general task T_{gen}. For example, if T_1 has the goal of turning ON a red button in location (1,2,3), and T_2 has the goal of turning ON a red button in location (4,5,6), then T_{gen} would have the goal of turning ON a red button without a location feature. When a feature is generalized from the goal, the system also tries to generalize the states in S_{task}, letting the task ignore that feature. Thus, T_{gen} can initiate in any location and any state with a red button ON achieves its goal.

This generalization is attempted each time a T_{new} is added to $Tasks$. If there exist two tasks T_1 and T_2 with similar goal states, then the system makes a general version of this task.

TABLE I

SOCIAL CUES FOR TRANSPARENCY IN A SOCIALLY GUIDED EXPLORATION

Context	Leo's Behavior	Intention
Human points to object	Looks at Object	Shows Object of Attention
Human present in workspace	Gaze follows human	Shows social engagement
Executing an Action	Looks at Object	Shows Object of Attention
Human says: "Look Leo, it's Task-X"	Subtle Head Nod and Happy facial	Confirms goal state of task-X
Human says: "Try to Act-Y the Obj-Z"	Look to human if suggestion is taken	Acknowledge partner's suggestion
Speech didn't parse, Unknown object request, Label without point	Confusion gesture	Communicates problem
Unconfident task execution	Glances to human more	Conveys uncertainty
Task is done, and human says: "Good!"	Nods head	Positive feedback for current option
Human asks a yes/no question	Nod/Shake	Communicates knowledge/ability
Intermittent	Eye blinks, Gaze shifts, Body motion	Conveys awareness and aliveness
Novel event	Surprise (raise brows/ears, open mouth)	Task being created.
Mastery triggers a task execution	Concentration (brows/ears down)	A known task is being tried
Completion of a successful task attempt	Happy (open mouth, raised ears)	Expectation met
Completion of a failed task attempt	Sad (closed mouth, ears down)	Expectation broken
Feedback from Human	Happy/Sad	Acknowledges feedback

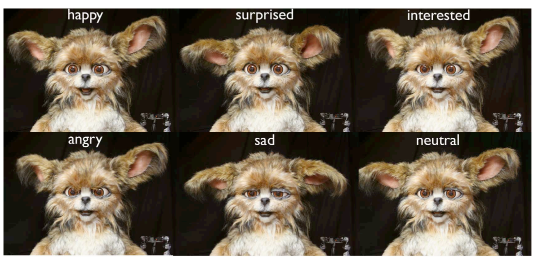

Fig. 5. Leo can use several facial poses to express internal learning state.

Two goals are similar if they differ by no more than four percepts. In generalizing S_{task} and G for all $T \in Tasks$, the generalization mechanism expands the portion of the state space in which tasks can be initiated or considered achieved. This results in an efficient representation, as the system continually makes the state space representations more compact. Additionally, it is a goal-oriented approach to domain transfer, as the system is continually refining the *context* and the *goal* aspects of the activity representation.

F. Transparency Mechanisms

Several expressive skills contribute to Leo's effectiveness as a social learner. Many are designed around theories of how humans communicate within a joint activity [5]. For example, principles of grounding. In all activity, humans look for evidence that their action has succeeded, and this extends to joint activity. Thus, the ability to establish a mutual belief that a joint activity has succeeded is fundamental to a successful collaborative activity.

Table I highlights many of the social cues that Leo uses to facilitate the collaborative activity of learning. Eye gaze establishes joint attention, reassuring the teacher that the robot is attending to the right object. Subtle nods acknowledge task stages, e.g., confirming when the teacher labels a task goal.

Additionally, Leo uses its face for subtle expressions about the learning state. The robot's facial expression shifts to a particular pose for fleeting moments (2-3 seconds), indicating an internal state and then returns to a neutral pose. The expressions are chosen to communicate information to the human partner, and are inspired by research showing that different facial action units communicate specific meanings [18] (Fig. 5). For example, raised eyebrows and wide eyes indicate heightened attention, which is the desired communication with Leo's surprised expression. This approach results in a dynamic and informative facial behavior.

Leonardo communicates various learning contexts to the human partner with facial expression (Table I). When the Novelty behavior is triggered, a fleeting surprised expression lets the human know that a task is being created. When the Mastery behavior causes a task to be practiced, Leo makes a concentrated facial expression and later a happy/sad expression upon the success/failure of the attempt. Throughout, if the human gives good or bad feedback, Leo makes a happy or sad expression to acknowledge this feedback. When the human labels a goal state Leonardo makes a happy expression and a head nod to acknowledge the labeling.

G. Scaffolding Mechanisms

Learning in a social environment is characterized by socially guided discovery, a balance between learning on one's own and benefiting from the social environment. The following are social scaffolding mechanisms at work on the Leonardo platform to enable Socially Guided Exploration.

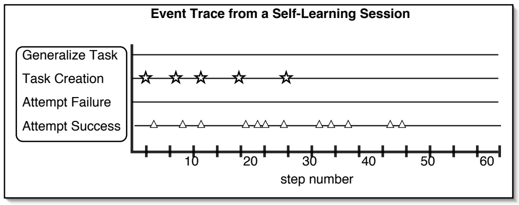

Fig. 6. Leo learning by himself.

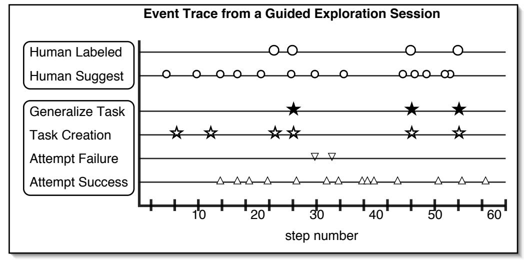

Fig. 7. Learning with a human partner.

Social attention: The attention of the robot is directed in ways that are intuitive for the human. Attention responds to socially salient stimuli and stimuli that are relevant to the current task. The robot tracks the pointing gestures and head pose of a human partner, which contribute to the saliency of objects and their likelihood for attention direction. For details on the robot's social attention system see [21].

Guidance: Throughout the interaction, the human can suggest actions for Leo to try. The human's request is treated as a suggestion rather than an interrupt. The suggestion increases the likelihood that the Explore behavior will trigger, but there is still some probability that Leo will decide to practice a relevant task or learn about a novel state change.

Recognizing goal states: Leo creates task representations of novelties in the environment. The human can facilitate this process by pointing out goal states with a variety of speech utterances (e.g., "Look Leo, it's X"). This serves to increase the likelihood that the Novelty behavior will trigger, creating a task with the label "X."

Environmental structure: A key contribution of the human teacher is their ability to physically structure the learning environment, highlighting salient elements. They draw the system into new generalizations, link old information to new situations, and point out when a learned task is relevant in the current situation.

IV. Resulting Learning Interactions

Socially Guided Exploration has been tested with a playroom scenario (Fig. 1).[3] Leo has toy boxes with three inputs (switches, levers, buttons), a lid and five LEDs. We can program the box's responses to actions on input devices (e.g., requiring an action sequence before the lid opens).

We have tested Socially Guided Exploration with several configurations of the playroom. Both simple tasks like changing the state of levers and switches, and more complex tasks like a sequence of actions to open the box lid. Here we detail one experimental setting and compare a session where Leo learns on his own to a session with a human partner.

In the experiment, Leo has four actions that can be directed toward objects: pressing down, sliding left/right, and a hand flip. Leo has no initial knowledge about the playroom, and uses the Socially Guided Exploration mechanism to build the $Tasks$ set about the various goal states in the playroom. We programmed the toy box with the following "Slide-Box" behavior: a press action toggles the color between green and red; when green the slide right action opens the lid, and the color must be red for a slide left action to close the lid. In this scenario, there are four general tasks/goals to learn: T_{open}–open the box, T_{closed}–close the box, T_{green}–make the box green, T_{red}–make the box red. Ideally, we would like the robot to have general versions of these tasks (e.g., T_{green} should not depend on the box being open/closed or on its location), but the robot will initially learn non-general versions and generalize its tasks with experience.

Figures 6 and 7 show events over the first 50 actions taken by the robot in the learning sessions. Figure 6 is from the robot learning by itself. Given a Slide-Box, the motivational drives trigger learning behaviors in a way to build a task set that eventually includes four tasks about the open, closed, green and red box states. The mastery drive triggers the practicing of these learned tasks throughout the session. However, this session was not long enough for Leo to learn generalizations that could be made in this scenario (e.g., it did not learn that the box can be made green regardless of the lid state; and the tasks are all location specific).

Figure 7, shows a session with a human partner in the same scenario.[4] We see that the human was able to successfully guide the robot's exploration by intermittently suggesting actions. Five of the six tasks created were states changes brought about by a human suggested action. Thus, the human helps shape the exploration and bring about interesting states. Additionally, the human partner was able to highlight interesting states, to make sure the robot did not miss a learning opportunity. Four of the six tasks created were pointed out by the human with a task label.

A final contribution of human partner was in helping the robot to generalize. In this experiment, due to the human's action suggestions the robot learned to generalize the T_{red} policy, learning that the box could be made red when the lid was open or closed (the first generalization in step 25). Additionally, the human could change the environment for

[3]All processes described in this paper run in real-time on a dual G5 Mac.

[4]The human subject (not one of the authors) was given a brief introduction with a list of utterances the robot understands and the box functionality.

the robot, moving the box to different locations. In the two later generalizations, the robot learns that the T_{red} and T_{open} policies should not depend on the location of the box. Thus, the human partner draws the robot into generalization opportunities through incrementally structuring the environment and making timely action suggestions.

V. Discussion

In designing robotic agents that learn new skills and tasks 'on the job' from everyday 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, etc.). This raises important research questions for the Machine Learning community. How do we design machines that learn effectively from human guidance? And, what is the right level of human interaction at a given time?

In prior works that incorporate a human into a machine learning process, the level of human interaction generally stays constant, remaining at one end of the guidance-exploration spectrum. Some are more guidance oriented, completely dependent on a human instruction. Others are more exploration based, using limited input from a teacher.

In this work, we recognize that a social learner needs both, and the Socially Guided Exploration mechanism brings these together in one learning system. Motivations drive exploration of the environment and the creation of goal-oriented tasks about novel events. A human partner can influence learning through: attention direction, action suggestions, labeling goal states, and positive/negative feedback.

An important feature of the Socially Guided Exploration mechanism is that the system frames its own learning problems. Many prior works where a machine learns a new task or skill assume that a goal is known (defined by the designer), is implicit in the given reward function, or the goal is to learn a world model. Alternatively, we address how a learner can be motivated to learn new tasks/goals, framing its own learning problem in a way that affords non-expert human guidance if available. A goal-oriented approach is fundamentally necessary for social learners, since their human partners implicitly interpret the world in intentional and goal-oriented ways.

VI. Conclusion

This work recognizes that a robot learning in a social environment needs the ability to both learn on its own and to take advantage of the social structure provided by a human partner. Our Socially Guided Exploration learning mechanism has motivations to explore its environment and is able to create goal-oriented task representations of novel events. Additionally this process can be influenced by a human partner through attention direction, action suggestion, labeling of goal states, and feedback. Thus, intrinsic measures along with extrinsic support bias behavior to create learning opportunities for a Reinforcement Learning mechanism.

References

[1] Christopher G. Atkeson and Stefan Schaal. Robot learning from demonstration. In *Proc. 14th International Conference on Machine Learning*, pages 12–20. Morgan Kaufmann, 1997.

[2] 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.

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

[4] C. Breazeal, M. Berlin, A. Brooks, J. Gray, and A. L. Thomaz. Using perspective taking to learn from ambiguous demonstrations. *to appear in the Journal of Robotics and Autonomous Systems Special Issue on Robot Programming by Demonstration*, 2005.

[5] H. H. Clark. *Using Language*. Cambridge University Press, Cambridge, 1996.

[6] J. Clouse and P. Utgoff. A teaching method for reinforcement learning. In *Proc. of the Nineth International Conf. on Machine Learning (ICML)*, pages 92–101, 1992.

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

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

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

[10] R. Maclin, J. Shavlik, L. Torrey, T. Walker, and E. Wild. Giving advice about preferred actions to reinforcement learners via knowledge-based kernel regression. In *Proceedings of the The Twentieth National Conference on Artificial Intelligence (AAAI)*, Pittsburgh, PA, July 2005.

[11] L-P. Morency, A. Rahimi, N. Checka, and T. Darrell. Fast stereo-based head tracking for interactive environment. In *Int. Conference on Automatic Face and Gesture Recognition*, 2002.

[12] M. N. Nicolescu and M. J. Matarić. Natural methods for robot task learning: Instructive demonstrations, generalization and practice. In *Proceedings of the 2nd Intl. Conf. AAMAS*, Melbourne, Australia, July 2003.

[13] P-Y. Oudeyer and F. Kaplan. Intelligent adaptive curiosity: a source of self-development. In *Proceedings of the 4th International Workshop on Epigenetic Robotics*, volume 117, pages 127–130, 2004.

[14] 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.

[15] J. Schmidhuber. Self-motivated development through rewards for predictor errors/improvements. In D. Blank and L. Meeden, editors, *Proc. Developmental Robotics 2005 AAAI Spring Symposium*, 2005.

[16] S. Singh, A. G. Barto, and N. Chentanez. Intrinsically motivated reinforcement learning. In *Proceedings of Advances in Neural Information Processing Systems 17 (NIPS)*, 2005.

[17] W. D. Smart and L. P. Kaelbling. Effective reinforcement learning for mobile robots. In *In Proceedings of the IEEE International Conference on Robotics and Automation*, pages 3404–3410, 2002.

[18] C. Smith and H. Scott. A componential approach to the meaning of facial expressions. In *The Psychology of Facial Expression*. Cambridge University Press, United Kingdom, 1997.

[19] R. Sutton, D. Precup, and S. Singh. Intra-option learning about temporally abstract actions. In *Proceedings of the Fifteenth International Conference on Machine Learning (ICML98)*, Masion, WI, 1998.

[20] R. Sutton, D. Precup, and S. Singh. Between mdps and semi-mdps: Learning, planning and representing knowledge at multiple temporal scales. *Journal of Artificial Intelligence Research*, 1:139, 1999.

[21] A. L. Thomaz, M. Berlin, and C. Breazeal. An embodied computational model of social referencing. In *IEEE International Workshop on Human Robot Interaction (RO-MAN)*, 2005.

[22] R. Voyles and P. Khosla. A multi-agent system for programming robotic agents by human demonstration. In *Proceedings of AI and Manufacturing Research Planning Workshop*, 1998.