

Learning from Human Teachers with Socially Guided Exploration

Cynthia Breazeal*

MIT Media Lab
20 Ames St., E15-468
Cambridge, MA 02139
cynthiab@media.mit.edu

Andrea L. Thomaz*

School of Interactive Computing
Georgia Institute of Technology
85 5th St. NW, Atlanta GA 30332
athomaz@cc.gatech.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 experiment with non-expert human subjects shows a human is able to shape the learning process through suggesting actions and drawing attention to goal states. Human guidance results in a task set that is significantly more focused and efficient, while self exploration results in a broader set.

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], [5] or by tutelage [10], [13]). 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, the human may control the reward signal given to a reinforcement learner [2], [8], [15], provide advice [7], [11], or control the agent’s actions during training [18]. These approaches have the benefit that learning does not require the human’s undivided attention. However, they strongly limit how the human can assist the learner and 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 (who are not familiar with Machine Learning techniques). Taking inspiration from human learning, however, we observe that a social learner must be able to move flexibly along this *guidance-exploration* spectrum. Children explore and learn on their own, but in the presence of a teacher they can take advantage of social cues and communication to accomplish more. Thus, we posit that a robot should be able to explore



Fig. 1. Leonardo in his workspace with a toy puzzle box.

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. Our experimental results highlight that the nature of what is learned in self-learning verses guided learning is different. We argue that robots will need to learn flexibly along the full guidance-exploration spectrum.

II. ROBOT PLATFORM

Our research platform 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 [12], 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, $O = \{o_1, \dots, o_k\}$. The Perception System is a set of *percepts* $P = \{p_1, \dots, p_n\}$. 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

This work was supported by the MIT Media Lab and Toyota Motor Corp.

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

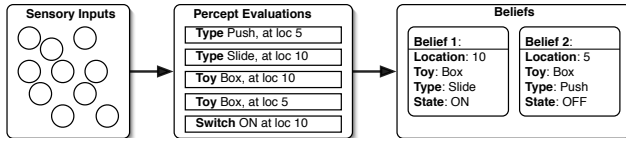


Fig. 2. Sensory input is classified by percepts and then merged into discrete object representations. In this timestep, 5 percepts yield 2 object beliefs.

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 $A = \{a_1, \dots, a_i\}$ be the set of Leo’s basic actions.

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* [9]). 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. These interactions help make the learning process more effective. This is the primary inspiration for our 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 [14], [16], [17]. 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

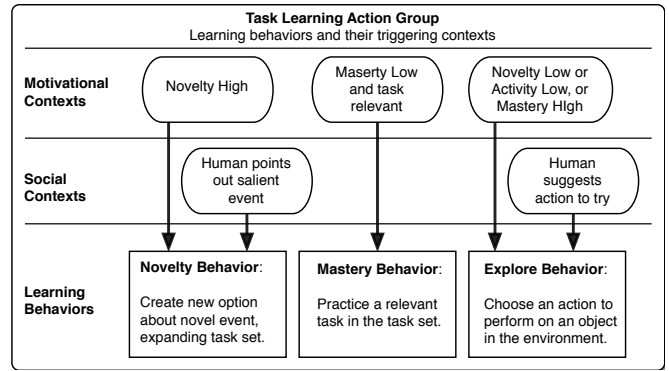


Fig. 3. The three learning behaviors and their social/motivational contexts.

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 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: $d_{chg}(s_1, s_2) = \frac{1}{\text{frequency}(s_1, s_2)}$. An event causes the drive to drift towards its max value for a period, $t = d_{chg}(s_1, s_2)t_{max}$. The maximum effect time, t_{max} , is 30 sec.

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 number of successful attempts over the total task attempts made.

B. Task and Goal Representation

Tasks and their goals are represented with *Task Option Policies*. This name reflects its similarity to the Options approach in Reinforcement Learning [21].

Each $T \in \text{Tasks}$ 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.

For simplicity, we refer to *Task Option Policies* as *tasks*.

Goals encode what must hold true to consider the task achieved. Specifically, a goal $G = \{x_1, \dots, x_y\}$ where every $x \in G$ represents a belief 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 expectation to).²

A task can be executed (is relevant) when the current state is in S_{task} . During execution, actions are chosen according to π' until the current state is in S_{goal} (with some probability of terminating early). A goal G is true (or is achieved) in a state s if: $\forall x \in G$, if any belief b in s matches all the *criteria* $\in x$, then b also matches all the *expectation* $\in x$.

C. 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-D and 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.

²This goal construct is also used in prior work, [4], [10].

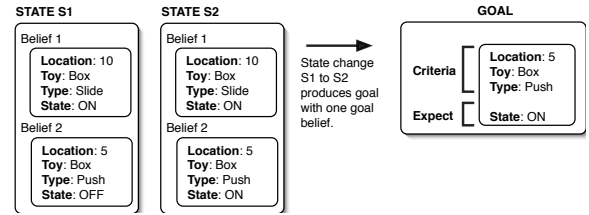


Fig. 4. A simple example of creating a goal from a state change.

D. Task Learning

These three behaviors result in a mechanism that learns object-oriented tasks. The Novelty behavior creates new tasks. 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., 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 [20]). 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. Two goals are similar if they differ by no more than four percepts. When a generalized task is created, the original more specific tasks are saved as part of the representation. In the future, we planned to use this as a way to let the system de-generalize if the abstraction did not prove useful. Additionally, any $T \in Tasks$ is removed from the set if it

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
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
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
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

has a very low confidence after several execution attempts (these thresholds are set empirically).

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

Leo has several expressive skills contributing to its effectiveness as a social learner. Many are based on theories of human joint activity [6]. 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 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 [19]. 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

The goal of our approach is for a robot learner to strike a balance between self learning and social learning. The following are social scaffolding mechanisms at work on the Leonardo platform to enable Socially Guided Exploration.

Social attention: The attention of the robot is directed in ways that are intuitive for the human, responding to socially salient stimuli and stimuli that are relevant to the current task. The robot tracks the human's pointing gestures and head pose, which contribute to the saliency of objects. For details on the robot's social attention system see [22].

Guidance: Throughout the interaction, the human can suggest actions for Leo to try. A 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 (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: An implicit 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. EVALUATION

To evaluate Socially Guided Exploration, we conducted an experiment with people interacting with the Leonardo robot.

A. Experimental Setup

To show experimentally which aspects of the learning process are influenced by a human partner's social scaffolding we collected data from two types of learning sessions:

- 1) SELF: The robot learning on its own.
- 2) GUIDED: The robot learning with a human teacher.

We collect several independent measures during the learning sessions. We then compare means between the two groups with t-tests³ in order to understand some specific details of how guidance and self learning differ.

Over the course of two days, we solicited 11 participants from the campus community (55% female). Due to corrupted log files, we only use 9 of these in the analysis. For the SELF condition, we collected data from 10 sessions of the Leonardo robot learning by itself in the same environment.

The experimental scenario is a playroom (Fig. 1), in which Leo has toy puzzle boxes. The boxes have three inputs (a switch, a slider, and a button), a lid, five colored LEDs, and can also output sounds. We can program how the box responds to actions on input devices (e.g., the actions required before the lid opens, or an LED turns on, etc.).

Leo has five primitive actions for the box (Button-Press, Slider-Left, Slider-Right, Switch-Left, Switch-Right), but no initial knowledge about the effects of these actions in the playroom. Leo uses the Socially Guided Exploration mechanism to build a *Tasks* set about the playroom. There is a single toy puzzle box programmed with the following input-output behavior: (1) Pressing the button toggles through the five LED colors: white, red, yellow, green, blue. (2) If both the slider and the switch are flipped to the left when the color is white, then the box lid opens. (3) If the slider and switch are flipped to the right when the color is yellow, then the box lid closes. (4) If the lid is open and the color changes to blue, then the box will play a song.

B. Instructions to Human Subjects

Subjects are shown the functionality of the puzzle box and told that their goal is to help Leo learn about it. They are told the robot is able to do some simple actions on the toy puzzle box, and once turned on, it will start exploring what it can do with the box. They are told they can help Leo by making action suggestions, by naming aspects of the box, and by testing that these named aspects have been learned. Finally, their goal in this interaction is to make sure that Leo learns to do three things: T_{Blue} —Make the light blue; T_{Open} —Make the lid open; T_{Song} —Make the song play. The experimenter shows subjects exactly how to achieve these tasks, and they are optionally given time to play with the box on their own before starting the learning session with Leo. They are told that Leo understands the following speech utterances:

- "Leo, try to...[press the button, move the slider left/right, move the switch left/right]"
- "Look Leo, It's...[Open, Closed, A Song, Blue, White, Green, Red, Yellow]"
- "Leo, Try to make it...[Open, Closed, Play a Song, Blue, White, Green, Red, Yellow]"

C. Results

None of the human teachers focused on the song task, thus these results focus on the tasks that people tried to teach,

³This assumes our data are approximately gaussian. We believe this is a common and fair assumption for the behavioral metrics we are analyzing.

T_{Blue} and T_{Open} . All 9 participants succeeded in getting the robot to reach the T_{Blue} and T_{Open} tasks. Everyone taught the T_{Blue} task first, and there was an average of 9 actions between first encountering the T_{Blue} and T_{Open} goals.

During the learning session, we log several measures to analyze the effect of guidance on the learning process. In addition to collecting metrics during the learning session, the efficacy of the learned task sets was tested in simulation afterwards (detailed below). The differences between the Self Exploration and Socially Guided Exploration cases are summarized in Table II. In all we tested 12 hypothesized differences between the guided and self learning groups. Here we detail the six significant differences found.

The human partner is able to guide the robot to the desired goal states faster than happens on its own. This is seen in the difference between groups in the number of actions to the first encounter of any of the three experiment goals. The average for GUIDE, 3.56, is significantly less than the average for the SELF condition, 11.2. Thus, people were able to utilize the social scaffolding mechanisms to focus the robot on aspects of the environment that they wanted it to learn about. This is also supported by qualities of the resulting *Tasks* set that is learned. In the GUIDE condition, the resulting *Tasks* sets were more related to the experiment goals (i.e., T_{Blue} , T_{Open} or T_{Song} is true in a task's goal state). We see a significant difference in both the number of tasks related to T_{Blue} and T_{Open} (see Table II).

Finally, the Socially Guided Exploration learns a better task set for achieving the experiment goals. In the post analysis of the learned tasks, we tested each task set from a test suite of five initial states, looking for their ability to achieve the experimental goals. Each experiment goal has a different test suite of five initial states, three of which are very close to the goal (1 or 2 actions required), two of which are farther away (more than 2 actions required). For each of the learned *Tasks* sets, we record the number of actions needed to reach each of the experimental task goals from each of the test states. We found some significant differences in the generality of the learned tasks. The average number of states that the GUIDE condition sets could reach the T_{Open} goal, 1.56, was significantly better than the average in the SELF condition, 0.58. And though we didn't find this difference for the T_{Blue} goal, we do see that the GUIDE condition is significantly faster at achieving T_{Blue} in the post analysis than the SELF condition, 1.69 versus 2.66. Thus, human guidance leads to task sets that are better at achieving the experimental goals at a later time.

V. DISCUSSION

In designing robotic agents that learn new skills and tasks 'on the job' from everyday people, we recognize that the typical person is not familiar with machine learning techniques, but is intimately familiar with various forms of social learning (e.g., tutelage, imitation, etc.). This raises important research questions. How do we design machines that learn effectively from human guidance? And, what is the right level of human interaction at a given time?

TABLE II

DIFFERENCES BETWEEN SELF AND SOCIALLY GUIDED EXPLORATION

Mesure	Mean SELF	Mean GUIDE	T-Test Results
# actions to reach first goal in learning session	11.2	3.56	$t(19) = 2.11$; $p < .05$
Size of resulting T_{asks} set	10.4	7.55	$t(19) = 7.18$; $p < .001$
# tasks for T_{Blue}	0.833	1.333	$t(19) = -2.58$; $p < .01$
# tasks for T_{Open}	1	1.77	$t(19) = -1.83$; $p < .05$
# Init States can reach T_{Open} in post-experiment	.58	1.56	$t(19) = -2.88$; $p < .01$
# actions to reach T_{Blue} in post-experiment	2.66	1.69	$t(19) = 2.19$; $p < .05$

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.

Our experiments show that Socially Guided Exploration is successful in allowing members of the general public to guide the robot's learning process. People were able to focus the robot's learning to particular goals that they desired. Furthermore, compared to self-learning in the same environment, the learning of these goals is accelerated and the resulting representation of these tasks is more useful at a later time. The task sets learned under guidance are smaller and more closely related to the specific tasks that people were trying to teach. In self-learning on the other hand, the robot learned a broader task set — serendipitously learning aspects of the environment that the human was not focused on teaching. While this knowledge about the environment may not be what the human had in mind today, it could certainly be advantageous in the future. Clearly both learning types are beneficial in different ways, supporting our approach of covering the full guidance-exploration spectrum.

VI. CONCLUSION

We posit that robots that learn in human environments need to successfully learn on their own *and* take advantage of the social structure provided by human teachers. Our Socially Guided Exploration learning mechanism motivates our robot to explore its environment 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

the robot's behavior to create learning opportunities for a Reinforcement Learning mechanism that can successfully leverage the presence of a human teacher to learn in a more efficient goal-directed manner, and also exploit intrinsically driven exploration and novelty to serendipitously learn new knowledge that can bootstrap future learning.

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] S. Chernova and M. Veloso. Confidence-based policy learning from demonstration using gaussian mixture models. In *Proc. of Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2007.
- [6] H. H. Clark. *Using Language*. Cambridge University Press, Cambridge, 1996.
- [7] J. Clouse and P. Utgoff. A teaching method for reinforcement learning. In *Proc. of the Ninth International Conf. on Machine Learning (ICML)*, pages 92–101, 1992.
- [8] F. Kaplan, P.-Y. Oudeyer, E. Kubinyi, and A. Miklosi. Robotic clicker training. *Robotics and Autonomous Systems*, 38(3-4):197–206, 2002.
- [9] Ed. M. Cole L. S. Vygotsky. *Mind in society: the development of higher psychological processes*. Harvard University Press, Cambridge, MA, 1978.
- [10] A. Lockerd and C. Breazeal. Tutelage and socially guided robot learning. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2004.
- [11] 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 (AAI)*, Pittsburgh, PA, 2005.
- [12] 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.
- [13] 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, 2003.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] S. Singh, A. G. Barto, and N. Chentanez. Intrinsically motivated reinforcement learning. In *Proceedings of Advances in Neural Information Processing Systems 17 (NIPS)*, 2005.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.