Exploiting Social Partners in Robot Learning

Maya Cakmak · Nick DePalma · Rosa I. Arriaga · Andrea L. Thomaz

June, 2010

Abstract Social learning in robotics has largely focused on imitation learning. Here we take a broader view and are interested in the multifaceted ways that a social partner can influence the learning process. We implement four social learning mechanisms on a robot: *stimulus enhancement*, *emulation*, *mimicking*, and *imitation*, and illustrate the computational benefits of each. In particular, we illustrate that some strategies are about directing the attention of the learner to objects and others are about actions. Taken together these strategies form a rich repertoire allowing social learners to use a social partner to greatly impact their learning process. We demonstrate these results in simulation and with physical robot 'playmates'.

Keywords Learning by Imitation · Social Learning · Biologically Inspired Learning

1 Introduction

Our work is inspired by the vision of service robots existing in human environment, assisting with various tasks in our homes, schools, and workplaces. Social learning will be crucial to the successful application of robots in everyday human environments. It will be impossible to give these machines all of the knowledge and skills a priori that they will need to serve useful long term roles in our dynamic world. The ability for everyday users, not experts, to guide them easily will be key to their success. Our research aims to build more flexible, efficient, and teachable robots, and is motivated by the distinction between human learning and machine learning. The research question we explore in this work is: *What are the best computational models to use in exploiting information provided by a social partner?*

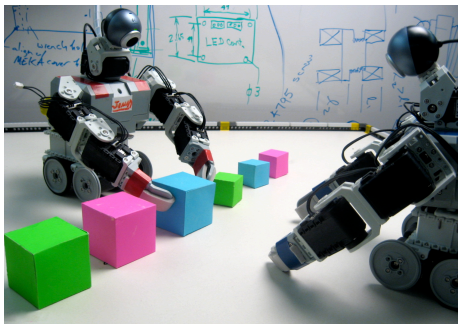


Fig. 1 Robot playmates Jimmy and Jenny in the playground.

1.1 Social Learning in Humans and Animals

Humans and some animals are equipped with various mechanisms that take advantage of social partners. Children naturally interact with adults and peers to learn new things in social situations. They are motivated learners that seek out and recognize learning partners and learning opportunities (Rogoff and Gardner, 1984; Pea, 1993), and throughout development, learning is aided in crucial ways by the structure and support of their environment and especially their social environment (L. S. Vygotsky, 1978; Greenfield, 1984; Lave and Wenger, 1991). These social partners guide a learn-

This work is supported by the National Science Foundation, award number IIS-0812106.

Center for Robotics and Intelligent Machines
College of Computing
Georgia Institute of Technology
801 Atlantic Drive
Atlanta, GA 30332-0760 USA
Phone 404.385.7480
Fax 404.894.0673

ing process in a variety of ways; for example, directing the learner's attention to informative parts of the environment (Wertsch et al., 1984; Zukow-Goldring et al., 2002). Understanding these mechanisms and their role in learning will be useful in building robots with similar abilities to benefit from other agents (humans or robots) in their environment, and explicit teaching attempts by these agents.

Our approach is motivated by the following four social learning mechanisms identified in natural learners (Tomasello, 2001; Call and Carpenter, 2002):

- *Stimulus (local) enhancement* is a mechanism through which an observer (child, novice) is drawn to objects others interact with. This facilitates learning by focusing the observer's exploration on interesting objects–ones useful to other social group members.
- *Emulation* is a process where the observer witnesses someone produce a particular result on an object, but then employs their own action repertoire to produce the result. Learning is facilitated both by attention direction to an object of interest and by observing the goal.
- *Mimicking* corresponds to the observer copying the actions of others without an appreciation of their purpose. The observer later comes to discover the effects of the action in various situations. Mimicking suggests, to the observer, actions that can produce useful results.
- *Imitation* refers to reproducing the actions of others to obtain the same results with the same goal.

Humans exhibit all of these social learning mechanisms, imitation being the most complex. While imitation seems to be distinctly human, many animals make use of the simpler relaxed versions of imitation. We are interested in taking a broad view of the ways that social partners influence learning. In particular, we believe the three relaxed versions of imitation learning are potentially useful in creating more natural social learning interactions between humans and robots.

1.2 Approach

In this article we show an implementation of these four social learning mechanisms and articulate the distinct computational benefits of each. One contribution of this work is our analysis of relaxed versions of imitation learning.

In order to directly compare these mechanisms we use a controlled learning environment, where the social partner is another robot. This allows us to systematically change the behavior of the social partner and understand the effect it has on the different learning strategies. We then draw conclusions about the computational benefits of each social learning strategy.

We show that all four social strategies provide learning benefits over self exploration, particularly when the target goal of learning is a rare occurrence in the environment. We characterize the differences between strategies, showing that the "best" one depends on both the nature of the problem space and the current behavior of the social partner. These results are demonstrated in simulation and with two physical robot 'playmates'.

1.3 Overview

In the following section we present related works, and in Section 3 we detail our implementation of the social learning mechanisms. Section 4 covers experiments with the baseline non-social learning strategies for comparison, and Section 5 is our social learning experiment and results. In Section 7 we consider several issues related to the generality of these findings (alternative performance metrics, effects of noise, and alternative classifiers). Finally we have a discussion of these results and their implications for future work in social robot learning in Section 8.

2 Related Work

In this section we briefly review some approaches to social learning in robots.

Several prior works deal with the scenario of a machine learning by observing human behavior. Learning high-level tasks by observation (Kuniyoshi et al., 1994; Voyles and Khosla, 1998), using a human demonstration to learn a reward function (Atkeson and Schaal, 1997), and skill learning by demonstration (Schaal, 1999; Breazeal and Scassellati, 2002). There is usually a specific training phase, where the machine observes the human, then a machine learning technique is used to abstract a model of the demonstrated skill.

In order to imitate, the robot has to map a sensed experience to a corresponding motor output. Many have focused on this perceptual-motor mapping problem. Often this is learned by observation, where the robot is given several observations of a particular motor action (Demiris and Hayes, 2002; Jenkins and Mataric', 2002; Alissandrakis et al., 2006).

In other works the human is able to directly influence the actions of the machine to provide it with an experience from which to learn. In one example, the robot learns a navigation task by following a human demonstrator who uses vocal cues to frame the learning (Nicolescu and Matarić, 2003). A related example

has a robot learn a symbolic high level task within a social dialog (Breazeal et al., 2004).

The pick and place method of programming is widespread in industrial robotics, allowing an operator to manipulate the robot and essentially record a desired motion trajectory to be played back. Calinon and Billard have looked at how a person could similarly demonstrate task examples to a robot by moving its arms, generalizing a motion trajectory representation for the task (Calinon and Billard, 2007). Others let a human directly control the actions of a robot agent with teleoperation to supervise a Reinforcement Learning (RL) process (Smart and Kaelbling, 2002), or to provide example task demonstrations (Peters and Campbell, 2003). Some recent approaches have the agent provide feedback about when these demonstrations are needed. In confidence-based learning (Chernova and Veloso, 2007), the robot requests additional demonstrations in states that are different from previous examples. In our own prior work (Thomaz and Breazeal, 2008; Thomaz and Cakmak, 2009b), the agent communicates uncertainty with eye gaze. Similarly in (Grollman and Jenkins, 2008), the robot communicates certainty in order to solicit demonstrations from the teacher.

In other cases the human influences the experience of the machine with higher level constructs than individual actions, for example, providing feedback to a reinforcement learner. Several approaches are inspired by animal training techniques like clicker training and shaping (Blumberg et al., 2002; Kaplan et al., 2002; Saksida et al., 1998). A human trainer uses instrumental conditioning techniques and signals the agent when a goal behavior has been achieved. Related to this, a common approach for incorporating human input to a reinforcement learner lets the human directly control the reward signal to the agent (Isbell et al., 2001; Stern et al., 1998). (Thomaz and Breazeal, 2008) have augmented this approach to interactive RL.

A few studies have also taken inspiration from non-imitative social learning mechanisms seen in animals and humans. Melo et al. (2007) present a framework for RL in which relaxed versions of imitation learning involve observing a subset of the information in an expert demonstration. For example, in a strategy that corresponds to *emulation*, the learner only observes the sequence of states during a demonstration (as opposed to a complete transition-reward tuple sequence). Lopes et al. (2009) present a computational model of social learning in which the behavior of the learner depends on a weighted sum of three sources of information: action preferences, observed effects and inferred goals. They show that different weight distributions result in behaviors that are similar to social learners in cited experiments with chimpanzees and children. One distinction of our work is our approach to modeling the various social learning strategies through changes in an attention mechanism. Additionally, our evaluation compares the computational benefits of each of the four strategies across various environments.

Any approach that takes input from a human teacher has to determine the amount of teacher involvement in the process. Prior work has investigated a wide spectrum of teacher involvement. From systems that are completely dependent on the teacher in order to learn anything, to others that do self-learning and incorporate some human feedback and guidance along the way.

One high level point we take away from social learning in humans and animals is the ability to flexibly operate along this spectrum of teacher engagement. The four social learning mechanisms we implement here represent different points on this spectrum: from imitation (complete dependence on the teacher's demonstration) to relaxed versions of imitation that are biased by the teacher in various ways. Our experiments with these mechanisms illustrate how these strategies are mutually beneficial and argue for a social learning approach that incorporates a variety of ways to exploit social partners.

3 Implementation

In this work, we have a social learning situation composed of two robot playmates with similar action and perception capabilities. Our experiments focus on learning a "sound-making" affordance for different objects in the environment.

We use two robots, Jimmy and Jenny (Fig. 1), which are upper torso humanoids on wheels built from Bioloid kits and Webcams. Their 8 degrees of freedom enable arm movements, torso rotation and neck tilt. The wheels are used to navigate the workspace.

The behavior system is implemented in C6, a branch of the latest revision of the *Creatures* architecture for interactive characters (Blumberg et al., 2002). This controls the real robots with percepts from sensors, as well as a graphical model of the robots with simulated sensing, world dynamics, and virtual objects. In simulation, we can set up environments composed of different object properties (Fig. 2).

The behavior system implements a finite state machine to control the exploration for collecting learning experiences. In non-social exploration the robot (i) observes the environment, (ii) approaches the most salient object, (iii) performs the selected action, (iv) observes the outcome (sound or no sound), (v) goes back to its initial position and (vi) updates the saliency of objects

Table 1 Features and feature values of objects.

Feature	# of Values	Values
Color	4	Pink, Blue, Green, Orange
Size	3	Small, Medium, Large
Shape	2	Cube, Sphere

Table 2 Sound-maker property of objects in different learning environments in simulation.

Description of sound-makers	Number of sound-makers	Percentage
All objects with color other than green	18	75%
All green and orange objects	12	50%
All green objects	6	25%
All green objects that are not large	4	∼17%
All small and green objects	2	∼8%
Only the small, green, cube-shaped object	1	∼4%

and actions based on its exploration strategy. In social exploration, after each object interaction the robot goes into an *observing* state and performs the same updates, of object saliency and action desirability, based on its observation of the other agent's interaction. In the rest of this section we give details on the domain of our simulation experiments and our implementation of exploration strategies. Details specific to the physical robot experiment are given in Section 6.

3.1 Objects

The learning environment involves objects with three discrete perceived attributes: *color*, *size* and *shape*, and one hidden property of *sound-maker* (see Table 1). In our experiments, the environment always contains all possible combinations of *color*, *size* and *shape*, however the *sound-making* properties of these objects can change. Different learning problems are obtained by changing the percentage of objects that make sound in the environment. For instance, all green objects could be sound makers in one environment, while in another, all objects with a particular shape and size are sound-makers.

Based on prior work (Thomaz and Cakmak, 2009a), we hypothesize that social learning will be especially beneficial in the case of rare sound-makers; thus, we systematically vary the frequency of sound-makers in the environment to compare various non-social and social exploration strategies.

The simulation environment has 24 objects with different attributes (one of 4 colors, 3 sizes and 2 shapes). We control the percentage of objects in the environment that produce sound, resulting in six learning environments as described in Table 2. These environments are chosen to cover a range of different learning problems where the target class varies from frequent to rare. Note that there are a number of environments with the same fraction of sound-makers and the choice of the particular environment used in our experiment is arbitrary. The choice of these particular fractions is also arbitrary, however it is intended to cover the range with more emphasis on rare sound-maker environments.

3.2 Perception

The social learning task considered in this study requires several perceptual capabilities:

1. Detecting objects in the environment and their properties
2. Detecting objects that are being interacted by a social agent
3. Detecting the actions performed by a social agent
4. Detecting the effects of own actions or social partner's actions on objects

All of these perceptual problems are trivialized in the simulation experiment by making all information available to the learner directly from the internal data structures of the simulator. Some of these problems are also simplified on the real robots by the fact that we are using two identical robots (with the same action repertoire and perceptual capabilities) and by constraining the environment.

3.3 Actions

The playmates' action set has two actions: *poke*–a single arm swing (*e.g.*, for pushing objects) and *grasp*–a coordinated swing of both arms. Both involve an initial *approach* to an object of interest, and are parametrized with the following discrete parameters (i) acting distances and (ii) grasp width or (iii) poking speed. In simulation we use 24 different actions (poke or grasp, 4 grasp widths, 4 poke speeds and 3 acting distances) as summarized in Table 3.

As with objects, we vary the frequency of sound-producing interactions by tuning the actions to have different effects on the objects, yielding different learning problems. This is achieved by making only one or both of the actions able produce sound and by varying the range of grasp width, poking speed and acting distance within which an action produces sound.

Table 3 Action parameters and their values.

Action	Parameter	# of Values	Values
	Action Type	2	Grasp, Poke
Grasp	Distance	3	Far, Middle, Close
	Width	4	Very-large, Large, Small, Very-small
Poke	Distance	3	Far, Middle, Close
	Speed	4	Very-fast, Fast, Slow, Very-slow

Table 4 Actions that produce sound in different learning environments in simulation.

Description of sound producing actions	# of such actions	Percentage
Both actions, except when (Grasp) *width* is *very-large*, or (Poke) *speed* is *very-slow*.	18	75%
Both actions, when *width* is *very-small, small*, or *speed* is *very-fast, fast*.	12	50%
Both actions, when *width* is *very-small*, or *speed* is *very-fast*.	6	25%
Both actions, when *distance* is not *far*, and *width* is *very-small*, or *speed* is *very-fast*.	4	∼17%
Both actions, when *distance* is *close*, and *width* is *very-small*, or *speed* is *very-fast*.	2	∼8%
Only grasp, when *distance* is *close*, and *width* is *very-small*.	1	∼4%

In the simulation experiments, we have six cases in which a different fraction of the action set is able to produce sound when executed on a sound-maker object (Table 4).

3.4 Learning Task

Our experiments focus on the task of learning a relation between a *context* in which an *action* produces a certain *outcome*; referred to as *affordance learning*. These relations are learned from interaction experiences of `[context-action-outcome]` tuples (Sahin et al., 2007). We use a 2-class Support Vector Machine (SVM) classifier (Vapnik, 1998) to predict an action's effect in a given environmental context. We consider other classifiers in Section 7.2.

SVMs are widely used discriminative classifiers. SVM classification is based on a linear separator constructed from a set of representative points close to the margin (support vectors). The input space is implicitly projected to a higher dimensional space with the help of kernels. A linear kernel was used in our experiments. For implementation, we use the open source library LibSVM (Chang and Lin, 2001) integrated with the WEKA data mining software (Hall et al., 2009).

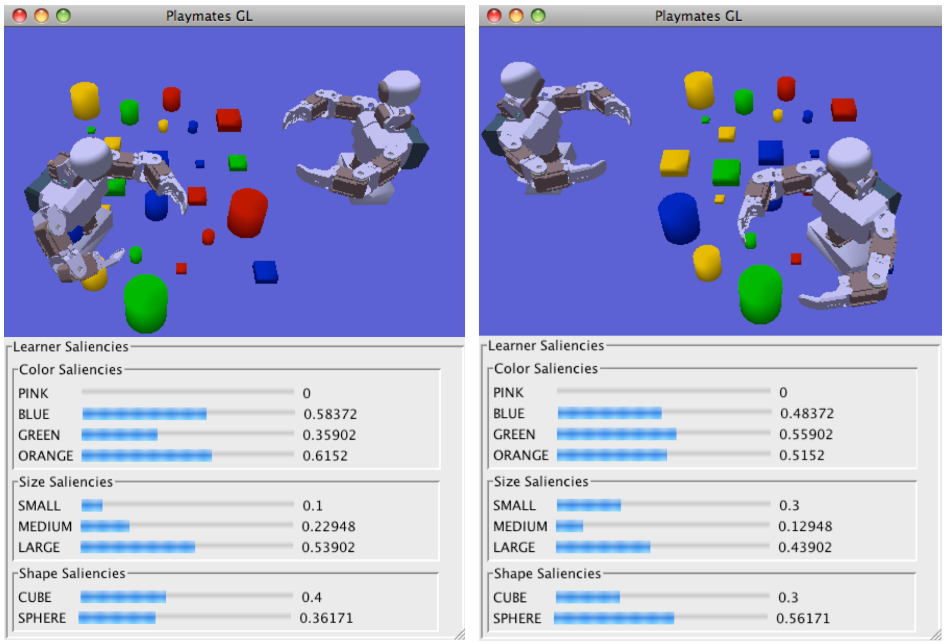
The SVM inputs are the values of perceived features of the interacted object (Table 1) and parameters of the action performed on that object (Table 3). The input vector contains one variable for each possible value of features and parameters. The variable corresponding to the current value is set to 1 while all other values are set to 0. The prediction target is whether or not this `context-action` produces sound. In this framework the robot is simultaneously learning the object features and action parameters required to produce a desired effect in the environment. Exploration is the process of collecting these interaction experiences.

3.5 Exploration Strategies

Our goal is to compare social and non-social *exploration strategies*, *i.e.* rules for interacting with the environment to collect data. The exploration strategy determines which object in the environment to interact with next, and what action to perform. Similar to the way that animals benefit from social learning mechanisms, these strategies can help a robot by guiding its exploration of the space of possible objects and actions towards useful parts of the space. While stimulus enhancement and emulation direct the learner's attention to informative parts of the *object space*, (*i.e.* the environment) *mimicking* guides the learner in the *action space*. *Imitation* combines the benefits of both types of information.

The alternative to social learning is non-social learning, in which a robot can use various exploration strategies. For instance, it can randomly select an object and try all possible actions on it, or it can adapt its exploration based on previous interactions.

In this study, an exploration strategy is implemented as an attention mechanism, where each object attribute value and action parameter value has a corresponding saliency ranging from 0 to 1. The robot always performs the most salient action on the most salient object. For example if the saliencies of the four possible values of the color attribute are as given in Fig. 2(b) (pink: 0.00, blue: 0.48 green: 0.56 orange: 0.51) the robot will interact with a *green* object because it has the highest saliency. Other feature values of the object are chosen in a similar way. As a result, given the saliency distribution in Fig. 2(b) the robot interacts with the *green*, *large*, *sphere* object.



(a) Social partner's object interaction (b) Learner's object interaction

Fig. 2 Snapshot of the simulation experiments in C6. The bars show the learner's object attribute saliencies for *color*, *size* and *shape*.

Action selection works similarly. Each action parameter value (e.g. *far, middle, close* for the Grasp-Distance parameter) has a corresponding saliency. There is an extra parameter that can take two values (grasp or poke) that determines which of the two actions is used in the interaction. The robot chooses the action with higher saliency and uses the parameter values with highest saliency for the parameters of the chosen action.

Object and action selection is the same in all exploration strategies (*i.e.* select most salient), whereas the way that saliencies are updated is different in all strategies. Each strategy has a different rule for updating saliencies after every interaction. Details on how each strategy updates saliencies is given in Sec. 4 and 5.

As an example, consider the exploration strategy used by the learner in Fig. 2. This strategy increases the saliency of the attributes of the object that the social partner interacts with and decreases the saliency of different object attributes. Since the social partner interacts with a *green* object (Fig. 2(a)), the saliency of *green* is increased by 0.2 while the saliency of other colors is decreased by 0.1. Similarly the saliency of *small* and *sphere* are increased. As a result of this update, the learner interacts with a similar object (*large, green, cube* as shown in Fig. 2(b)). Note that even though the saliency of *large* was increased and the saliency of *small* was decreased in the update, the saliency of *small* was large enough in the previous interaction to make the learner interact with a small object once more. Note also that the saliencies of actions and action parameters are randomized in this strategy, therefore it can be considered as a strategy that guides the learner in the *object space*.

We remark that this implementation of exploration strategies is *feature-based* rather than *object-based*. While this provides a simple mechanism to explore objects based on feature similarities, it is limited in terms of recognizing object identity.

3.6 Experimental Method

We conducted a series of experiments in order to analyze and compare social and non-social exploration strategies. In each experiment the learner uses a particular strategy to collect a data set which is used for training a sound-maker classifier. The experiment is repeated in several environments that present different learning problems. In social learning experiments the social partner has one of three pre-defined behaviors that are described in Sec. 5.2.

Different environments have different frequencies of sound producing interactions. Whether or not rareness is due to object or action has a different impact on the object-oriented versus action-oriented social learning strategies. Thus, we systematically experiment with both kinds of rareness. In these experiments, we first keep the percentage of sound producing actions constant at 25% and vary the sound-maker object rareness;

and then keep the object percentage constant at 25% and vary the percentage of sound producing actions.

We present experimental results from simulation for all environments, exploration strategies and social partner behaviors. Then we present a validation of these experiments with physical robots. In simulation there are 576 (24x24) possible test cases (interactions). SVM classifiers are trained in a batch mode after 28 interactions. This corresponds to a small subset of all possible interactions (∼5%). For each environmental condition, the experiments are repeated 200 times with random initialization. We report average performance across these repeated experiments.

With classifiers trained using the various strategies, we can compare performance. In doing this comparison, our performance measure is recall rate in prediction of the effect for all `object-action` combinations. Recall corresponds to the ratio of true positives and the sum of true positives and false negatives (i.e., of all the sound-making cases in the test set, the percentage predicted as such). In a later section (Sec. 7) we consider alternate performance metrics.

4 Baseline Experiments: Non-social Learning

An initial question for this research is the selection of a fair or appropriate non-social learning baseline to compare social learning against. We consider three different non-social exploration strategies for learning affordances: *random*, *goal-directed* and *novelty-based* exploration. We also compare these strategies with a *systematic* data set that consists of all possible interactions. This section describes the exploration strategies and experimental results for non-social learning. The details of how saliencies are updated for individual exploration strategies are summarized in Table 6.

4.1 Implementation of Exploration Strategies

(1) Random: In each interaction the robot randomly picks a new object, action and a set of action parameters. This is achieved by randomizing the saliency of each object attribute and action parameter and selecting the most salient object and action. The data sets collected with random exploration are equivalent to random subsets of the systematic data set.

(2) Goal-directed: In goal-directed exploration, the robot keeps interacting with objects similar to ones that have given the desired effect in a previous interaction. Likewise, it performs actions that are similar to those that produced sound in the past. If an interaction produces sound, the saliency of some attribute values of the object used in that interaction are increased and the saliency of different ones are decreased. Increasing or decreasing all attributes deterministically is avoided because this will result in interacting with the exact same object once it has produced sound, therefore will stop the exploration. By updating a random subset of the attributes of an object that made sound, the robot will interact with objects that have common attributes, rather than exactly the same object.

In the goal-directed strategy, if no sound is produced the robot updates saliencies randomly. As a result, the robot only pays attention to positive information. An alternative strategy could reduce the saliency of attribute values of the object used in an interaction that did not produce sound, in order to avoid objects similar to the ones that do not make sound.

(3)Novelty-based: In this strategy the robot prefers novel objects and actions. After every interaction the robot reduces the saliency of attribute values of the object that it just interacted with, while increasing the saliency of different values. Actions and action parameters are altered similarly.

(4) Systematic Data Set: In addition to the data sets collected using three exploration strategies, we consider a training set that consists of all possible object-action pairs in the learning space. Note that the data sets obtained with the exploration strategies have much fewer examples than the systematic dataset. In fact, the number of examples collected with exploration strategies is chosen to be a small fraction of all possible examples (*e.g.* in simulation 28 interactions is 5% of the size of the systematic data set).

4.2 Results for Non-social Learning

Our experiments let us compare the performance of the three individual exploration strategies, showing that novelty-based exploration performs best. Additionally we look at the effect of rareness of sound-makers in the environment, and number of interactions allowed.

4.2.1 Comparison of strategies

The systematic training set is designed to cover the complete learning space. Training with the systematic data set is a best-case scenario for the learning algorithm; it demonstrates how well the affordances can be learned when complete and equally distributed data is available and essentially shows that this is a learnable problem.

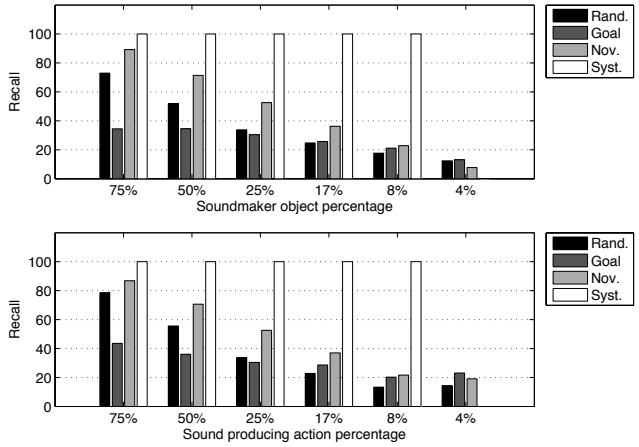


Fig. 3 Recall rate for non-social strategies after 28 interactions for (a) six environments with different sound-maker frequencies (sound producing action frequency held constant at 25%) and (b) six environments with different sound-producing action frequencies (sound-maker object frequency held constant at 25%).

A 20-fold cross validation test is performed on the systematic data set for 12 environments with varying degrees of sound-maker object and action rareness. We observe that prediction is 100% accurate for the systematic strategy in all environments with sound-maker frequency of 8% or greater (Fig. 3).

In the last environment case (4% sound-makers) the event of sound-making happens so infrequently that the resulting SVM always predicts 'no sound' and the recall rate is 0%. Standard SVMs are known to have a bias towards the larger class in the case of unbalanced datasets (Huang and Du, 2005). In this case (4% sound-maker objects and 25% sound producing actions, or the other way around) the systematic dataset is highly unbalanced: it has 6 positive interactions (interaction that produced sound) out of a total of 576 interactions.

Fig. 3 compares the recall rate for non-social learning strategies in different environments. The performance of the random exploration strategy reduces as the sound-maker objects become rare in the environment, since it is less likely to randomly interact with a sound-maker when it is rare.

The novelty-based strategy outperforms the other exploration strategies especially when the sound-makers are frequent. The strength of this strategy in these environments is its uniform coverage of the search space by always interacting with different objects. As the sound-makers become very rare the performance of all three strategies degrade and the difference between the strategies becomes less significant.

The goal-directed strategy results in lower recall rates than random when the sound-makers are frequent in the environment. With this strategy the robot interacts only with a subset of objects that are similar to the first object that was discovered to be a sound-maker. However, when the environment has a high percentage of sound-makers, objects with no common perceptual attributes may also be sound-makers. Therefore, in such environments covering only a subset of objects degrades the performance of the goal-directed strategy. As the sound-makers become less frequent the goal-directed strategy becomes better than the random strategy.

In the last environment, we observe that all strategies have non-zero recall unlike the systematic dataset. Even though the sound producing interactions happen rarely, the resulting data sets are less unbalanced since they include only a total of 28 interactions. As result the average recall rate is non-zero.

Table 5 Effect of (a) sound-maker object rareness and (b) sound producing action rareness on different exploration strategies (measured with 1-way ANOVA). These tests show that performance is significantly different as the target becomes rare with all of the strategies.

(a)

Strategy	Analysis of variance
Random	$F(5, 1194) = 93.91$, $p < .001$
Goal-directed	$F(5, 1194) = 11.51$, $p < .001$
Novelty-based	$F(5, 1194) = 178.19$, $p < .001$

(b)

Strategy	Analysis of variance
Random	$F(5, 1194) = 138.20$, $p < .001$
Goal-directed	$F(5, 1194) = 10.51$, $p < .001$
Novelty-based	$F(5, 1194) = 130.99$, $p < .001$

4.2.2 Comparison of environments

In Fig. 3, we see a significant effect of the rareness of sound-makers in the environment on all three exploration strategies (see Table 5 for statistical significance). While the performance of random and novelty-based strategies monotonically decrease with decreasing sound-maker frequency, the performance of the goal directed strategy increases initially and decreases afterwards for reasons explained above.

4.2.3 Comparison of number of interactions

All three strategies result in imperfect learning because they cannot explore the complete object/action space. However, we expect that the longer we allow the robot to interact with the environment, the better its learning will be. In Fig. 4 we present learning curves for non-social exploration strategies in three sample environments. In general, it can be observed that learning improves with increasing number of interactions. However, in the case of very rare sound-makers (4%), increasing the number of interactions does not improve

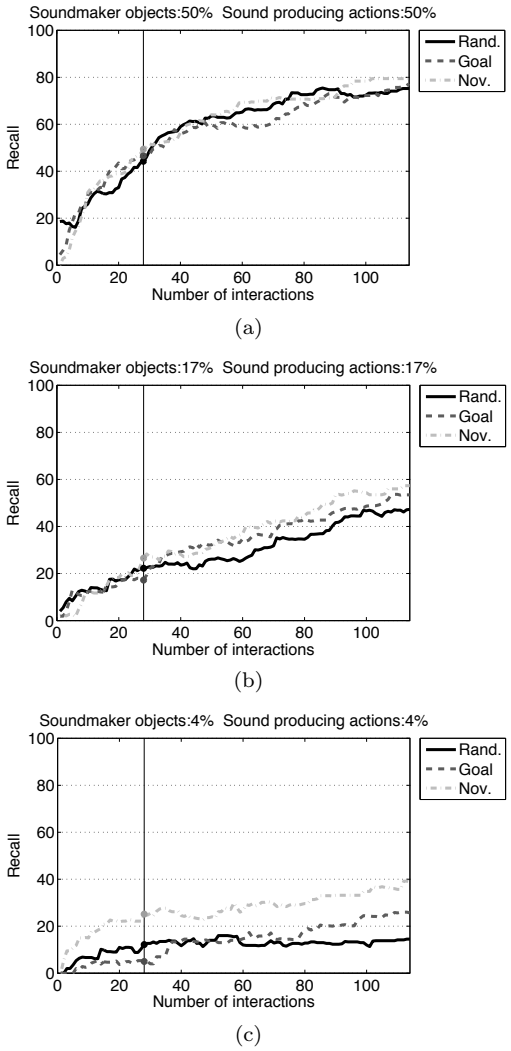


Fig. 4 Learning curves (*i.e.* change of recall rate with increasing number of interactions) for non-social learning methods in three sample environments with different action and object space rareness.

performance. The reason being that when the soundmakers are very rare, even 116 interactions is often not sufficient to randomly discover a sound-maker.

5 Experiments with Social Learning

Next, we present experiments evaluating social exploration strategies. In these experiments one of the robots (*learner*) explores the environment using a social strategy while the other robot (*social partner*) has a predefined behavior. The social partner behavior influences how much the learner can benefit from the social partner as a 'teacher', therefore we systematically vary it. We first present the implementation of the strategies and social partner behaviors, followed by comparative results.

5.1 Implementation of Social Exploration Strategies

As with the non-social strategies, the four social exploration strategies are implemented by varying the ways in which object and action saliencies are updated after each interaction with the environment.

(1) Stimulus Enhancement: The robot prefers to interact with objects that its playmate has interacted with. After every observed interaction, the learner increases the saliency of attributes of the object that the social partner has interacted with and decreases others.

(2) Emulation: The robot prefers objects seen to have given the desired effect. If an observed interaction produces sound, the saliencies of the attributes of the object used are increased. Otherwise, the saliencies are randomized.

(3) Mimicking: This strategy involves copying the actions of the social partner. We implement two versions:

- *Blind*: The learner mimics every partner action.
- *Goal-based*: The learner mimics actions only after it observes the goal.

Use of the term 'mimicking' in animal behavior literature is closer to *blind*, but this distinction is useful in illustrating computational differences between the social mechanisms.

(4) Imitation: In imitation, the learner focuses on the objects used by its social partner and copies the actions of the social partner. Again, there are two versions:

- *Blind*: The learner always imitates its social partner.
- *Goal-based*: It imitates after it observes the goal.

Both stimulus enhancement and emulation influence object attribute saliencies, but do not imply anything about actions. Action selection is random in these strategies. On the other hand, mimicking influences action saliencies while having no implication on objects. Object saliencies are updated randomly in mimicking. Imitation combines the strength of both, varying both the object and action saliencies based on the observation of the social partner. The social exploration strategies and their use of object, action and result components of the demonstration are summarized in Fig. 5. The implementation details for saliency update rules are given in Table 6.

Table 6 Saliency update rules of all exploration strategies. f denotes an object feature (*e.g.* color), v_f denotes an object feature value (*e.g.* green), $sal(v_f)$ denotes the saliency of the feature value, p_a denotes parameters related to action a (including the *ActionType* parameter), p denotes any action parameter, v_p denotes the value of an action parameter (*e.g.* very-far), $sound_{prev}$ denotes whether the learner's previous interaction produced sound, $sound_{partner}$ denotes whether the interaction of the social partner produced sound. Note that saliency values are bounded to the $[0, 1]$ range to avoid divergence. The upper and lower bounds are asserted after every update.

Strategy	Update Rules	
	Object Saliencies	Action Saliencies
Random	$\forall f : \forall v_f : sal(v_f) \leftarrow rand(0,1)$	$\forall p : \forall v_p : sal(v_p) \leftarrow rand(0,1)$
Goal-directed	**if**$(sound_{prev})$ $\forall f$: $\forall v_f = v_{f_{prev}} : sal(v_f) \leftarrow sal(v_f) + 0.2$ $\forall v_f \neq v_{f_{prev}} : sal(v_f) \leftarrow sal(v_f) - 0.1$ **else**: $\forall f : \forall v_f : sal(v_f) \leftarrow rand(0,1)$	**if**$(sound_{prev})$ $\forall p_{a_{prev}}$: $\forall v_p = v_{p_{prev}} : sal(v_p) \leftarrow sal(v_p) + 0.2$ $\forall v_p \neq v_{p_{prev}} : sal(v_p) \leftarrow sal(v_p) - 0.1$ **else**: $\forall p : \forall v_p : sal(v_p) \leftarrow rand(0,1)$
Novelty-based	$\forall f : \forall v_f = v_{f_{prev}} : sal(v_f) \leftarrow sal(v_f) - 0.1$ $\forall v_f \neq v_{f_{prev}} : sal(v_f) \leftarrow sal(v_f) + 0.2$	$\forall p_{a_{prev}} : \forall v_p = v_{p_{prev}} : sal(v_p) \leftarrow sal(v_p) - 0.1$ $\forall v_p \neq v_{p_{prev}} : sal(v_p) \leftarrow sal(v_p) + 0.2$
Stimulus enh.	$\forall f : \forall v_f = v_{f_{partner}} : sal(v_f) \leftarrow sal(v_f) + 0.2$ $\forall v_f \neq v_{f_{partner}} : sal(v_f) \leftarrow sal(v_f) - 0.1$	Same as Random.
Emulation	**if**$(sound_{parner})$ $\forall f$: $\forall v_f = v_{f_{partner}} : sal(v_f) \leftarrow sal(v_f) + 0.2$ $\forall v_f \neq v_{f_{partner}} : sal(v_f) \leftarrow sal(v_f) - 0.1$ **else**: $\forall f : \forall v_f : sal(v_f) \leftarrow rand(0,1)$	Same as Random.
B. Mimicking	Same as Random.	$\forall p_{a_{partner}} : \forall v_p = v_{p_{partner}} : sal(v_p) \leftarrow sal(v_p) + 0.2$ $\forall v_p \neq v_{p_{partner}} : sal(v_p) \leftarrow sal(v_p) - 0.1$
G. Mimicking	Same as Random.	**if**$(sound_{parner})$ $\forall p_{a_{partner}}$: $\forall v_p = v_{p_{partner}} : sal(v_p) \leftarrow sal(v_p) + 0.2$ $\forall v_p \neq v_{p_{partner}} : sal(v_p) \leftarrow sal(v_p) - 0.1$ **else**: $\forall p : \forall v_p : sal(v_p) \leftarrow rand(0,1)$
B. Imitation	Same as Stim. Enhancement.	Same as B. Mimicking.
G. Imitation	Same as Emulation.	Same as G. Mimicking.

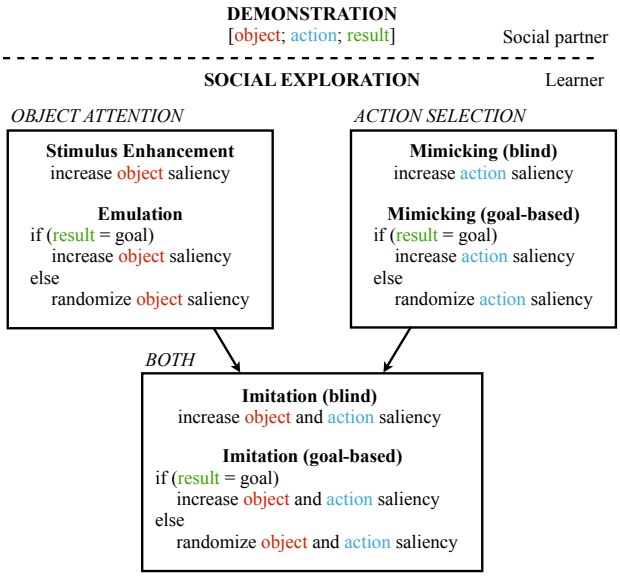


Fig. 5 Implementation of the social learning mechanisms and their use of object, action and result information from the social partner's demonstration.

5.2 Social Partner Behavior

The behavior of the social partner has a crucial effect on the learner. With particular social partner behaviors, these exploration strategies can become equivalent. For instance if the partner produces a sound with every interaction, stimulus enhancement and emulation behave very similarly. If the partner explores objects and actions randomly, a learner that blindly imitates will learn as if it was exploring randomly itself. Therefore to compare the strategies fairly, we systematically vary the behavior of the social partner.

There are four possible types of demonstrations in terms of the useful information communicated to the learner:

- *Goal-demonstration*: The target goal (sound) is shown with an appropriate action (sound-producing action) and appropriate object (sound-maker object).
- *Action-demonstration*: A sound-producing action is demonstrated on a non-sound-maker object.
- *Object-demonstration*: A non-sound-producing action is performed on a sound-maker object.
- *Negative-demonstration*: A non-sound-producing action is performed on a non-sound-maker object.

Social partner behaviors emerge as a result of different demonstration preferences. We consider three behaviors, summarized in Table 7:

Table 7 Demonstration type preferences for three social partner behaviors.

Demo. Type	Same-goal	Different-goal	Focused-demo.
Goal-demo.	60%	20%	20%
Action-demo.	20%	20%	80/0%
Object-demo.	20%	20%	0/80%
Neg.-demo.	0%	40%	0%

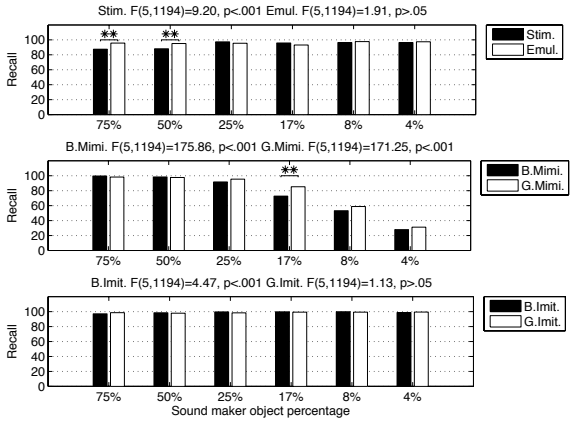
Social partner with ***same goal****:* In this case, the goal of the social partner largely overlaps with that of the learner. The partner spends a lot of time demonstrating the goal.

Social partner with ***different goal****:* Here, the goal of the partner has a small overlap with the learner and it spends little time demonstrating the goal. We do not define a particular goal for this social partner but we assume that it is related to an effect that is different from *sound* and that as a result the *sound* effect is demonstrated infrequently.

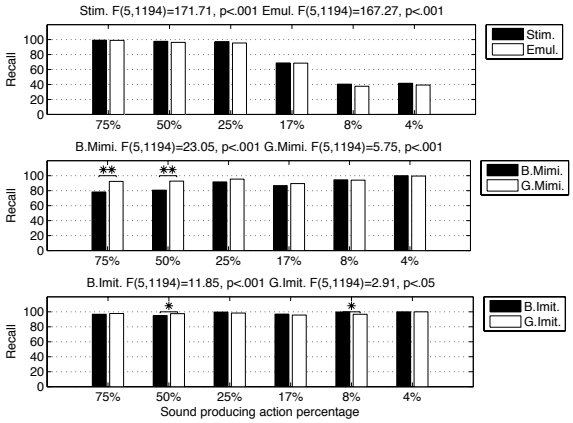
Social partner with ***focused demonstration****:* In the third case, the partner spends most of its time focusing either on the target action or object, without producing the goal. Focused demonstration can be considered a typical kind of teaching behavior. A teacher who is trying to teach a particular action might demonstrate it on an arbitrary object. Similarly, a teacher might present objects that are known to have useful affordances to the learner but let the learner discover what actions produce the desired effects on the object.

From the perspective of the learning strategies these different social partners give different amounts of information. While the *same goal* partner gives useful information most of the time for all strategies, the *different goal* partner rarely gives useful information. The social partner with *focused demonstrations* also gives useful information all the time, but it is partial. Strategies that pay attention to the wrong part of their demonstrations, or strategies that pay attention only when the goal is observed will not benefit from such partially useful demonstrations.

These social partner behaviors are ones we believe are fairly generic, and will transfer well to a situation in which a human is the social partner. Experiments with human trainers is left as future work. The goal of this work is to show the computational differences between strategies in a controlled learning environment.



(a)



(b)

Fig. 6 Comparison of social learning mechanisms for (a) different sound-maker object frequencies and (b) different sound producing action frequencies. The social partner has the *same goal* as the learner.

5.3 Results of Social Learning Experiments

As in the non-social experiments, the different environments correspond to different frequencies of sound producing objects or actions, which we systematically vary. In this section we present results from simulation for all environments, strategies and social partner behaviors.

Performance of social learning with a *same goal* social partner is presented in Fig. 6 for environments with different sound-maker object frequencies and different sound producing action frequencies. Similarly, performance for learning with a *different goal* social partner is given in Fig. 7; and for learning with a *focused demonstrations* social partner is given in Fig. 8. In this section, we analyze these results with respect to the environments in which each strategy is preferable. The effect of sound-maker rareness on learning performance, as determined by one-way ANOVA tests, are reported on

each graph. Additionally, the significance level of the difference between the two strategies plotted in each graph according to a T-test are indicated (* for $p < .05$, ** for $p < .005$). The T-tests indicate the difference between the blind and goal-directed versions of the strategies that focus on a particular aspect of the learning space (object space, action space or both).

5.3.1 Comparison with non-social exploration

Comparing Fig. 3 and Fig. 6 we observe that social learning usually outperforms non-social learning. However, when the learned affordance is not rare, random and novelty-based exploration have comparably high performance. Non-social learning in such cases has two advantages: (1) it does not require social partners and (2) it is less perceptually demanding on the learner in terms of identifying social partners and perceiving actions performed and objects used.

Additionally, non-social strategies can do better in environments with high sound-maker frequency when they are allowed to interact for a longer duration. For instance doubling the number of training interactions raises the performance of random and novelty-based exploration to 90-100% in environments with 75% and 50% sound-makers (Cakmak et al., 2009). Fig. 4 also shows that increasing the number of interactions improves the performance for non-social exploration especially in environments with frequent sound-makers. Since there's no requirement of a social partner, it's acceptable to perform non-social exploration for longer durations to collect more interaction samples.

5.3.2 Paying attention to objects

As observed in Fig. 6(a), increasing object rareness does not affect the performance of object focused strategies (stimulus enhancement and emulation) but it significantly reduces the performance of action focused strategies (mimicking). This suggests that when the object with the desired affordance is very rare, it is useful to let the social partner point it out. By randomly exploring actions on the right object the learner can discover affordances.

5.3.3 Paying attention to actions

Similarly, when the sound producing actions are rare, doing the right action becomes crucial. Performance of mimicking stays high over reducing sound-producing action frequencies (Fig. 6(b)).

Practically, mimicking will often be more powerful than the object-focused strategies since action spaces are usually larger than object spaces (which are naturally restricted by the environment). For instance, the most salient feature combination may be large-red-square, but if there is no such object the robot may end up choosing small-red-square. Generally, all feature combinations are not available in the environment, but all actions are. In these experiments, action and object spaces had the same number of possible configurations, thus having similar rareness effects.

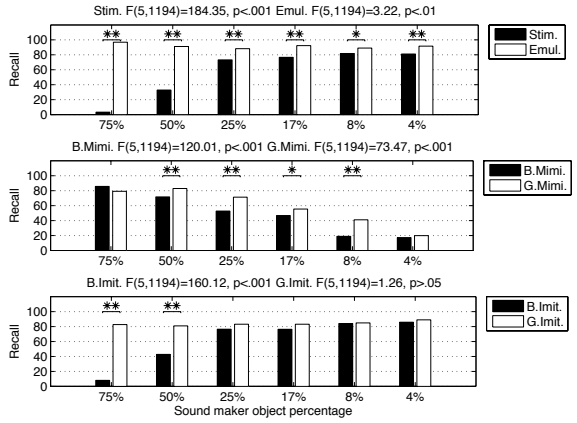
5.3.4 Imitation

Following from the previous two cases when everything is rare the most powerful strategy is imitation. As observed from Fig. 6 imitation performs well in all environments. This raises a question as to why imitation should not be the default strategy. There are two main disadvantages to always using imitation. First, it is the most computationally demanding for the learner; it requires paying attention to the context and and the action. Second it is also demanding of the demonstrator. For instance in the case where sound-maker objects are rare but the sound producing action is not, the demonstrator can just perform an *object-demonstration* rather than a *goal-demonstration* (Sec. 5.2). A robot could be equipped with other means for directing the attention of the learner to the right object without a demonstration. Examples include pointing to the object, pushing the object towards the learner, shaking the object, gazing at the object or putting away all other objects.

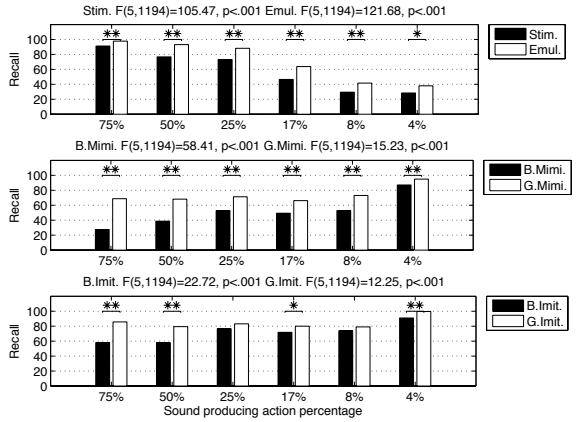
5.3.5 Paying attention to the goal

The performance of stimulus enhancement and emulation are very similar in Fig. 6. Likewise there are few significant differences between goal-based and blind strategies for mimicking or imitation. This suggests that when interacting with a social partner with the same goal as the learner, paying attention to the effect of demonstrations is less important. The attention of the learner is already attracted to the object that was interacted with, which happened to also produce sound since a high fraction of the demonstrations do so.

If the social partner has a different goal we observe that the performance of blind strategies is lower than that of goal-based strategies as shown in Fig. 7. In this case, blindly copying aspects of the demonstration results in an exploration focused on the wrong objects or actions. In other words they are misled to uninformative parts of the context and action spaces. Goal-based strategies, on the other hand, only pay attention to the social partner's useful demonstrations. The rest of the

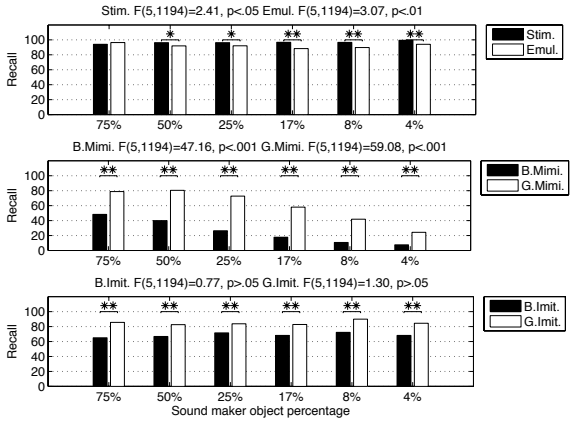


(a)

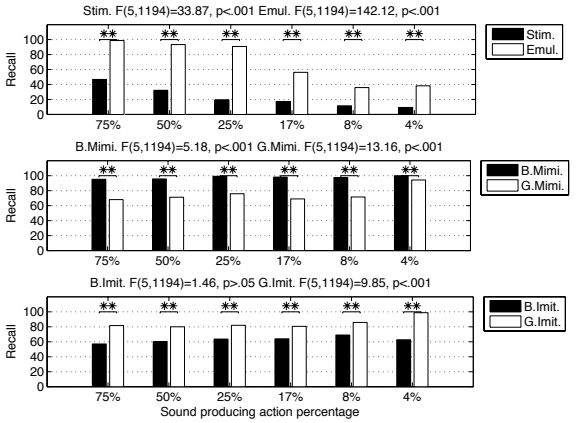


(b)

Fig. 7 Comparison of social learning mechanisms for (a) different sound-maker object frequencies and (b) different sound producing action frequencies; social partner has a *different goal*.



(a)



(b)

Fig. 8 Comparison of social learning mechanisms for (a) different sound-maker object frequencies when the social partner demonstrates the sound-maker objects and (b) different sound producing action frequencies when the social partner demonstrates the sound producing actions without actually producing sound–*focused demonstration*

time they randomly explore based on this useful information and thus have a higher chance to discover and gain experience with sound-makers.

In Fig. 8 we observe that the performance of the blind strategies are better than those of the goal-based strategies when the social partner performs focused demonstrations of objects or actions without producing the desired effect. The blind strategies benefit from these demonstrations by being directed to the right parts of their action or context space while the goal-based strategies ignore these demonstrations.

We could imagine a teacher providing action demonstrations on an arbitrary object, or simply presenting objects that are known to have useful affordances to the learner. In such cases it is useful to trust the teacher even if the goal has not been observed. By trusting the teacher the learner later comes to uncover the use of copied actions.

5.4 Asymmetry between object and action spaces

It can be noticed that the performance in similar parts of the object and action space are not exactly symmetric for similar behaviors. For instance in Fig. 7, in (a) at high sound-maker object frequencies the performance of emulation is as high as 90-100%, whereas in (b) at high sound-producing action frequencies the performance of goal-directed mimicking is about 70%. This is due to a subtle difference between the representation of object and action spaces. The action space consists of two independent smaller subspaces corresponding to each action. Learning about the parameters of one action does not provide any information for the other action and therefore both actions need to be explored sufficiently. For instance if the robot is performing a grasp, the values of poking parameters are meaningless. Additionally

the robot needs to simultaneously learn which action is useful in a given situation, as well as its parameters. On the other hand interaction with one object provides information about all attributes in the object space since all objects are represented with a value for each attribute. This makes the action space harder to explore than the object space. As a result the performance of object focused strategies in the object space, is better than the performance of action focused strategies in the action space.

6 Validation on the Physical Robots

A simplified version of the simulation experiments were run on the physical robots as described in Section 5. In this section we first give implementation and experimental details specific to the physical experiment and then present results that support our findings from the simulation experiment.

6.1 Implementation

Objects and Actions: As noted earlier, in practice the action space is often much larger than the object space. Accordingly, in the physical experiments we have 4 objects (2 colors, 2 sizes) and 18 possible actions (poke or grasp, 3 grasp widths, 3 poke speeds and 3 acting distances).

Perception: In the real robot experiment the configuration of objects in the environment is assumed to be fixed. When the learner decides to interact with a specific object it first navigates to the a known location that is approximately in front of the desired object. Then it uses the location of the object in the camera image and the known neck angle to adjust its distance to the object. The objects are detected in the camera by filtering the image for pre-defined color templates and finding connected components (blobs) in the filtered image. Features of the desired object (color and size) are also verified based on the detected blob. Sound detection is based on pitch thresholding through the microphones embedded on the webcams. The sound-maker objects have coins inside which makes them produce a detectable sound when dropped or tapped.

Perception of the social partner is also simplified in the physical robot experiment. After each action of the social partner, the information about the object that was interacted with, the action that was performed and the outcome of the interaction (sound/no sound) is sent to the learner by its social partner.

Table 8 Recall rate in physical robot experiments.

Environment	Stim. Enh.	B. Mimicking
Act.:50%, Obj.:50%	70%	100%
Act.:50%, Obj.:25%	86%	60%
Act.:3%, Obj.:50%	0%	100%
Act.:3%, Obj.:25%	20%	100%

6.2 Experiments

The experiments are performed in two different environments where (i) all small objects make sound (50%) and (ii) only the small green object makes sound (25%). We consider only two cases in which (i) poke always produces a sound (50%) and (ii) only one particular set of parameters for the grasp produces a sound and poke does not produce a sound (∼3%). Thus there are four different learning settings with different combinations of action and object sound-making properties.

In the physical experiment there are 72 (4x18) possible test cases (interactions). SVM classifiers are trained in a batch mode 8 interactions (∼10% of all possible interactions). For each environmental condition, the experiments are repeated 5 times with random initialization. We report average performance across these repeated experiments.

As we focus on the effect of asymmetry between the object and action spaces in the physical experiment, we experimented with two social strategies that focus on the objects and actions respectively: stimulus enhancement and blind mimicking. The social partner in these experiments always demonstrates the goal.

6.3 Results

Table 8 gives the results of learning in four different environments for the two strategies. The given results are the averages over 5 runs of 8 world interactions. The results support our findings from simulation that the performance of stimulus enhancement is less affected by decreasing sound-maker percentage, while the performance of mimicking is less affected by the decreasing sound-producing action frequency. Furthermore, due to the asymmetry in the action and object spaces in the physical experiments, we observe that the reduction in the performance of mimicking is less severe.

7 Generality of Results

In this section we address several follow-up items related to the generality of our results. We identify the limitations of the results presented so far and provide

extensions for these results in a sample setting. First we consider alternative metrics to recall rate (accuracy and correct decision making), second we consider different classifiers for learning, and finally we look at the effects of noise on our results.

7.1 Alternate Performance Metrics

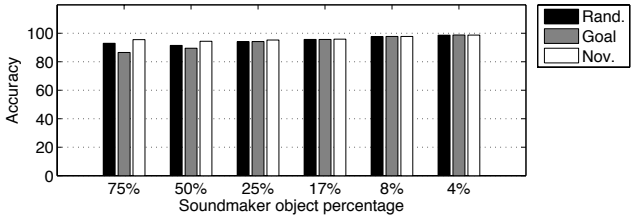
All of the previous results report recall rate. This corresponds to the correct prediction on the set of positive examples (i.e. examples in which a sound was produced as a result of the interaction). Although we motivate our choice of recall rate as the main performance metric in Section 7.1.2, this metric has some limitations. First, in cases where positive examples are very rare, recall rate reflects a test on very few samples. Secondly, this metric does not give any information about what the classifier will predict for negative interactions. In these respects, another metric of interest is accuracy (i.e., correct prediction on the complete systematic data set).

7.1.1 Accuracy

We look at accuracy results for both non-social and social strategies, across environments with different sound-maker object rareness. In Fig. 9, we see that accuracy remains high or increases for non-social strategies even though we know from our previous analysis (see Fig. 3) that recall rate goes down for all three strategies. When positive interactions are rare, the number of positive samples in the data set collected with non-social exploration has very few or no positive samples. This results in an overly negative classifier with a low recall rate. On the other hand, the accuracy of the classifier is very high since the test set mostly has negative samples, and the classifier predicts them correctly.

In the social strategies, we primarily see accuracy results that are complimentary to the recall results. For example, Fig. 10(b) shows the accuracy of the various social strategies, with a same-goal social partner, as sound-maker objects become more rare. When sound-makers are rare, we see that stimulus enhancement and emulation have consistent accuracy results (which agrees with the recall results). But the accuracy of mimicking goes up with rareness (which is the inverse of its falling recall performance). This is due to the classifier's propensity to predict false, which becomes more accurate as positive examples become rare.

In both of these examples, the accuracy metric agrees with the recall rate conclusions about social exploration strategies. When sound-maker objects are rare, stimulus enhancement or emulation would be preferred for



(a)

Fig. 9 Comparison of accuracy for non-social learning mechanisms on different sound-maker object frequencies.

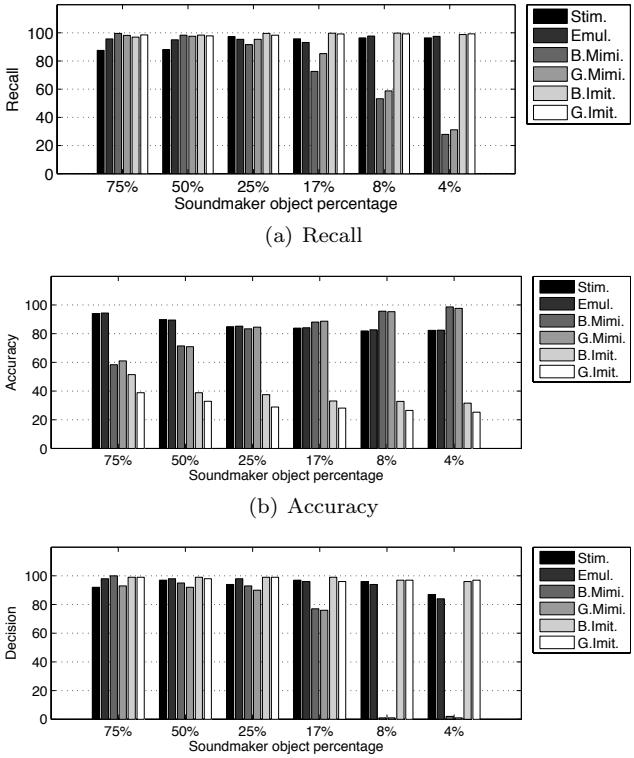
their high recall rate and relatively high accuracy. Similarly when sound producing actions are rare mimicking is preferable. What accuracy tells us in these cases is that the classifier is able to correctly predict negative cases as well as positive cases. This is a result of a data set with balanced positive and negative samples. With imitation on the other hand, we see that accuracy is quite low. Remember that imitation had excellent performance in the previous experiments, very high recall rates across various environments. However, the accuracy results show that the classifiers built from the imitation exploration strategies are too positive, and haven't had enough experience with the negative space to build a good model.

In the accuracy analysis of social strategies presented above, the social partner has the *same goal* behavior. Similar results can be seen for the *different goal* social partner, with accuracy results complimentary to recall results in every respect except for imitation. In the case of a different goal partner, goal-based imitation has a low accuracy (overly positive) classifier, whereas blind imitation does get some experience with negative examples and therefore has a better accuracy.

7.1.2 Correct Decisions

Since our end-goal is for the robot to be able to make appropriate use of its learned models, another performance metric we consider is the percentage of "correct decisions" the robot can make with its classifier. After learning we ask the robot to "make sound" and measure how often it can successfully choose an action-object combination to do so.

In some ways this is the most interesting metric for a robot learner, and it tests the applicability of the learned classifiers. For example, a classifier that always predicts 'no sound' may often be correct in terms of prediction, but it would not be able to produce sound because it does not know about any action-object pairs that make sound. On the other hand, a classifier that does predict 'sound' in some cases can decide to perform



(a) Recall

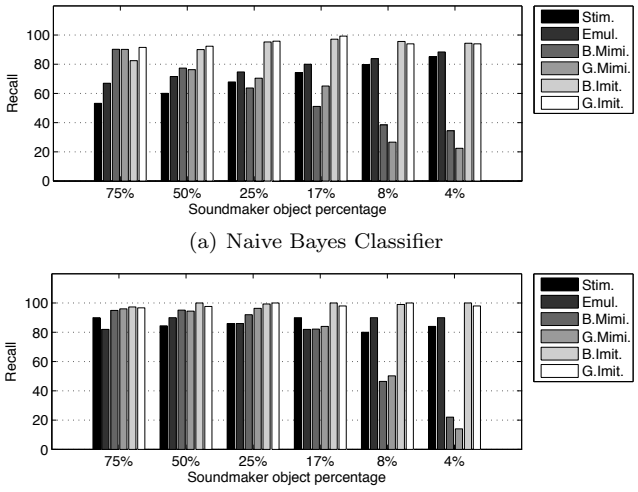
(b) Accuracy

(c) Correct Decision

Fig. 10 Comparison of social and non-social learning mechanisms for (a) different sound-maker object frequencies and (b) different sound producing action frequencies. The social partner has the *same goal* as the learner.

the interaction that predicts sound with the highest confidence. To measure correct decision performance, we do 100 randomly initialized learning sessions, and determine how many result in the ability to choose an action-object pair that makes sound.

When we look at this metric across all of the environments, exploration strategies and social partner behaviors, we see that it directly correlates with the recall results. As one example, Fig. 10(c) shows correct decision performance in environments with decreasing sound-maker object percentage for social learning strategies with a common-goal social partner. Similarly for other environments and social partner behaviors, correct decision percentage correlates to the recall rate. What this says is that even when learning results in a classifier that is too optimistic, the confidence about actually positive samples will be higher, so the robot will make correct decisions by choosing the interaction for which it is most confident that sound will be produced. When recall rate is low (that means the classifier recalls fewer or maybe zero of the available positive interactions) the chance of making a correct decision will



(a) Naive Bayes Classifier

(b) Decision Tree

Fig. 11 Comparison of different classifiers.

be lower or zero. Thus recall rate indicates how good the resulting decision making will be.

7.2 Different Classifiers

The results presented in this paper are based on the performance of an SVM classifier trained with the data collected with different exploration strategies in different environments. In order to investigate the limitations of this particular choice of classifier we analyze the learning performance of two classifiers other than SVMs: Naive Bayes classifiers (Langley et al., 1992) and Decision Trees. The *Naive Bayes classifier* learns the class-conditional probabilities $P(X{=}x_i|C{=}c_j)$ of each variable x_i (object features and action parameters) given the class label c_j (sound or no sound). It then uses Bayes' Rule to compute the probability of each class given the values of all variables and predicts the more probable class. A *decision tree* is a tree structure that describes the prediction process based on the values of each variable. Leaves on the tree correspond to predictions (sound or no sound) and branches correspond to different values or ranges of variables (object features or action parameters). The decision trees are learned using the C4.5 algorithm (Quinlan, 1993). We use the WEKA (Hall et al., 2009) implementation of both classifiers with default parameters.

We trained these classifiers using data obtained with the different social exploration strategies. We focused on the particular situation where the social partner has a *common goal* and we present results for different sound-maker object frequencies while the sound producing property of actions is kept constant. The recall

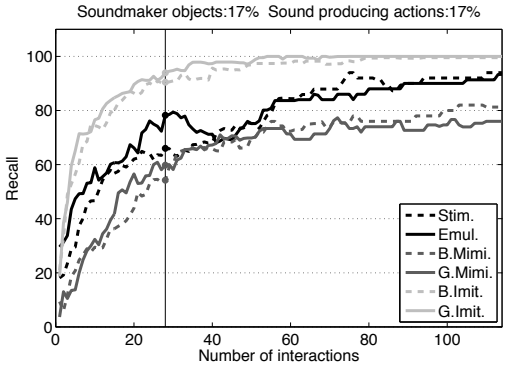
rates are shown on Fig. 11 and can be compared with the performance of SVM given in Fig. 10(a).

We observe that the performance of both classifiers are similar to SVMs, where we see the action focused strategies drop in recall performance as the as soundmaker objects become more rare. The performance of the Naive Bayes Classifier takes longer to converge, so it has slightly lower performance in all cases. This difference is more emphasized for Stimulus Enhancement and Emulation in environments with high percentage of sound-makers, which may be due to several reasons. First, these strategies get exposed to less positive examples as the social partner is not very helpful in these environments. Secondly, in these environments the positive class (sound) is determined by disjunctive variables (*e.g.* "all green and orange objects", Table 2) which takes more examples to learn since our input representation is essentially conjunctive (*i.e.* when an object is green it is also not-orange). The recall rate of Decision Trees is similar to SVMs in all environments. Even though no assertions can be made for other classifiers or learning algorithms based on these results, we demonstrate that three very different classifiers have very similar performances when using the same exploration strategy to collect data. The exploration strategies are independent of the learning algorithms. They effect the content of the data and how good the data set represents the concept that the classifier is learning. Thus, we can expect similar effects on other classifiers.

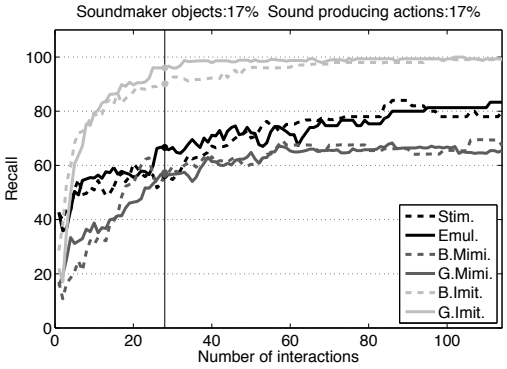
7.3 Effects of Noisy Data

In all of the simulation experiments presented here, the data was noise-free. Thus a final analysis we can look at is the effect of different amounts of noise in sound detection. For example, 5% noise means that randomly 5% of the interactions will do the opposite of what they were supposed to: if they were supposed to produce sound they will not, otherwise they will.

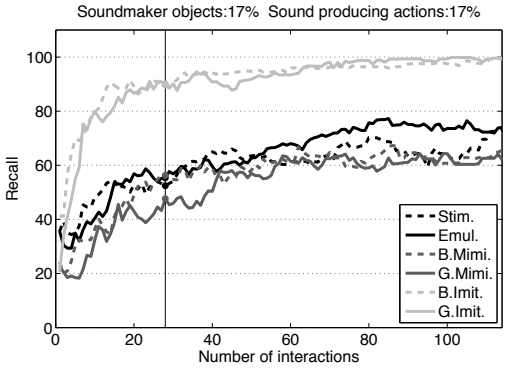
We analyze the effect of noise on learning curves. Fig. 12 shows the effect of 5%, 10% and 20% noise on social learning mechanisms for a fixed environment (sound-maker percentage: 17%, sound-producing action percentage: 17%, social partner behavior: common-goal). We observe that increasing noise reduces the convergence rate of the classifiers. As a consequence, the performance of the classifier at a given instance (*e.g.* after 28 interactions) is reduced. Similar results are obtained for different environments and social partner behaviors. Although the overall performance is degraded by noise, the observations made in the previous experiments holds for noisy data. In other words, the impact of noise is similar on all strategies and their relative performance remains the same in the presence of noise.



(a) 5% Noise



(b) 10% Noise



(c) 20% Noise

Fig. 12 Effect of different amounts of noise on learning curves for social learning strategies.

8 Discussion

In looking at social versus non-social learning, we see that social learning often out performs self-learning and is particularly beneficial when the target of learning is rare (which confirms prior work (Thomaz and Cakmak, 2009a)). However, when the target is relatively easy to find in the world, then many of the self-learning strate-

gies had performance on par with social exploration. One conclusion we draw from this work is that rather than using one or the other, self and social learning mechanisms will likely be mutually beneficial within a single robot learning framework.

The bulk of our results in this work center on the computational benefits of four biologically inspired social learning mechanisms: stimulus enhancement, emulation, mimicking, and imitation. We demonstrated that each strategy leads to different learning performance in different environments. Specifically, we investigated two dimensions of the environment: 1) the rarity of the learning target, and 2) the behavior of the social partner. Furthermore, when the learning target is a rare occurrence in the environment this can be due to the size of the object (feature) space or the size of the action space, and we differentiated between these two in our analyses.

When the rareness of the target is due to the particular object space, then the mechanisms related to object saliency (stimulus enhancement, emulation, and imitation) perform best. These three all do equally well if the social partner is actively demonstrating the goal. However, if the partner is demonstrating other goals, or only one aspect of the goal (either action or object), then emulation and goal-based imitation outperform stimulus enhancement because they pay attention to the effects of the partner's action to ignore demonstrations unrelated to the target learning goal.

Alternatively, in an environment where only a few specific actions produce the goal, then action oriented mechanisms (mimicking and imitation) are best. Again, when the social partner is demonstrating the goal, both do equally well. Otherwise, goal-based mimicking and imitation are preferred as they pay attention to effects.

Perhaps not surprisingly, goal-based imitation is robust across all the test environments. This might lead one to conclude that it is best to just equip a robot learner with the imitation strategy. However, there are a number of reasons that a social robot learner should also consider making use of non-imitative strategies.

Imitation is not always possible. When the agents have different morphologies or different action repertoires, then the learner may not be able to copy the exact action of the teacher. In this case emulation is a good strategy, in which the learner tries to achieve the demonstrated effects using its own action set. Paying attention to effects as opposed to blind copying will also be beneficial when there are multiple goals the robot wants to learn. Various social partners might share only a subset of these target goals, hence only a portion of their demonstrations will be useful to the robot. Finally, requiring full demonstrations of the learning target may be a burden for the teacher, particularly when this teacher is a human partner.

Imitation does not take full advantage of a social partner. The learner should be able to make use of full demonstrations when available, but as our results have shown, social learning is more than just demonstrations. Using non-imitative mechanisms in conjunction with imitation learning can let a robot use more of the partner's input, taking advantage of their presence and interactions in the environment even when they are not actively giving demonstrations.

Imitation has a positive bias. When we compare the strategies based on accuracy instead of recall rate, we find that imitation has poor accuracy. Having seen a very positive and small swath of the problem space, imitation results in a classifier that is too optimistic. The imitation strategy is best suited for a learning algorithm that is not affected by a data set that is largely biased in the positive direction. Alternatively, this positive bias could work well within a framework of self and social learning in which self exploration accumulates negative examples to create a balanced data set.

Thus, it is not surprising that nature endows humans and animals with a variety of mechanisms for taking advantage of social partners in the learning process. Our computational analysis finds that each serve a different purpose in the learning process, and have benefits over the others depending on the environment (the rareness of the learning goal and the behavior of the social partner).

Inspired by biological systems, we conclude that to best take advantage of a social environment robots need a repertoire of social learning mechanisms. In our future work we are building a framework in which all four of the mechanisms presented here can operate simultaneously in conjunction with self-learning. The challenge becomes appropriately switching between strategies. A naïve approach is to adopt a new strategy when the current one ceases to be informative. A more sophisticated approach might look for social or environmental "cues" that indicate what "kind" of social partner is present and how to best take advantage of their interactions in the world.

9 Conclusion

We presented a series of experiments on four social learning mechanisms: stimulus enhancement, emulation, mimicking, and imitation. We looked at the task of a robot learning a sound-making affordance of different objects, while another robot (a social partner) interacts with the same objects. The contribution of this

work is the articulation of the computational benefit of these four social learning strategies for a robot learner. The fact that each strategy has benefits over others in different situations indicates the importance of a social learner having a repertoire of strategies available to take advantage of social partners.

References

Alissandrakis, A., Nehaniv, C., and Dautenhahn, K. (2006). Action, state and effect metrics for robot imitation. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication*.

Atkeson, C. G. and Schaal, S. (1997). Robot learning from demonstration. In *Proc. 14th International Conference on Machine Learning*, pages 12–20. Morgan Kaufmann.

Blumberg, B., Downie, M., Ivanov, Y., Berlin, M., Johnson, M., and Tomlinson, B. (2002). Integrated learning for interactive synthetic characters. In *Proc. of the ACM SIGGRAPH*.

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

Breazeal, C. and Scassellati, B. (2002). Robots that imitate humans. *Trends in Cognitive Science*, 6(11).

Cakmak, M., DePalma, N., Arriaga, R., and Thomaz, A. (2009). Computational benefits of social learning mechanisms: Stimulus enhancement and emulation. In *IEEE Intl. Conference on Development and Learning*.

Calinon, S. and Billard, A. (2007). What is the teacher's role in robot programming by demonstration? - Toward benchmarks for improved learning. *Interaction Studies. Special Issue on Psychological Benchmarks in Human-Robot Interaction*, 8(3).

Call, J. and Carpenter, M. (2002). Three sources of information in social learning. In Dautenhahn, K. and Nehaniv, C., editors, *Imitation in animals and artifacts*. MIT Press.

Chang, C.-C. and Lin, C.-J. (2001). *LIBSVM: a library for support vector machines*.

Chernova, S. and Veloso, M. (2007). Confidence-based policy learning from demonstration using gaussian mixture models. In *Proc. of Autonomous Agents and Multi-Agent Systems (AAMAS)*.

Demiris, J. and Hayes, G. (2002). Imitation as a dual-route process featuring predictive and learning components: a biologically plausible computational model. In Dautenhahn, K. and Nehaniv, C. L., editors, *Imitation in Animals and Artifacts*. MIT Press, Cambridge.

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

Grollman, D. H. and Jenkins, O. C. (2008). Sparse incremental learning for interactive robot control policy estimation. In *In IEEE International Conference on Robotics and Automation*.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. (2009). The weka data mining software: An update. *SIGKDD Explorations*, 11(1).

Huang, Y. and Du, S. (2005). Weighted support vector machines for classification with uneven training class sizes. In *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, pages 4365–4369.

Isbell, C., Shelton, C., Kearns, M., Singh, S., and Stone, P. (2001). Cobot: A social reinforcement learning agent. *5th Intern. Conf. on Autonomous Agents*.

Jenkins, O. C. and Mataric', M. (2002). Deriving action and behavior primitives from human motion data. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2002)*, pages 2551–2556.

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

Kuniyoshi, Y., Inaba, M., and Inoue, H. (1994). Learning by watching: Extracting reusable task knowledge from visual observation of human performance. *IEEE Transactions on Robotics and Automation*, 10:799–822.

L. S. Vygotsky, E. M. C. (1978). *Mind in society: the development of higher psychological processes*. Harvard University Press, Cambridge, MA.

Langley, P., Iba, W., and Thompson, K. (1992). An analysis of bayesian classifier. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 223–228.

Lave, J. and Wenger, E. (1991). *Situated learning: legitimate peripheral participation*. Cambridge University Press, Cambridge.

Lopes, M., Melo, F., Kenward, B., and Santos-Victor, J. (2009). A computational model of social-learning mechanisms. *Adaptive Behavior*.

Melo, F., Lopes, M., Santos-Victor, J., and Ribeiro, M. (2007). A unified framework for imitation-like behaviors. In *4th International Symposium in Imitation in Animals and Artifacts*.

Nicolescu, M. N. and Matarić, M. J. (2003). Natural methods for robot task learning: Instructive demonstrations, generalization and practice. In *Proc. of the 2nd Intl. Conf. AAMAS*.

Pea (1993). Practices of distributed intelligence and designs for education. In Salomon, G., editor, *Distributed cognitions: Psychological and educational considerations*. Cambridge University Press, New York.

Peters, R. A. and Campbell, C. L. (2003). Robonaut task learning through teleoperation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, Taipei, Taiwan.

Quinlan, R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA.

Rogoff, B. and Gardner, H. (1984). Adult guidance of cognitive development. In Rogoff, B. and Lave, J., editors, *Everyday cognition: its development in social context*. Harvard University Press, Cambridge, MA.

Sahin, E., Cakmak, M., Dogar, M., Ugur, E., and Ucoluk, G. (2007). To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior*, 15(4):447–472.

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

Schaal, S. (1999). Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*, 3:233242.

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

Stern, A., Frank, A., and Resner, B. (1998). 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. ACM Press.

Thomaz, A. and Cakmak, M. (2009a). Learning about objects with human teachers. In *HRI '09: Proceedings of the ACM/IEEE Intl. Conference on Human-Robot Interaction*.

Thomaz, A. L. and Breazeal, C. (2008). Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence Journal*, 172:716–737.

Thomaz, A. L. and Cakmak, M. (2009b). Learning about objects with human teachers. In *Proc. of the International Conference on Human-Robot Interaction (HRI); accept rate: 19%*.

Tomasello, M. (2001). *The Cultural Origins of Human Cognition*. Harvard University Press.

Vapnik, V. (1998). *Statistical Learning Theory*. Wiley, New York, NY.

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

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

Zukow-Goldring, P., Arbib, M. A., and Oztop, E. (2002). Language and the mirror system: A perception/action based approach to cognitive development. *Working draft*.