# From Manipulation to Communicative Gesture

Shichao Ou and Rod Grupen
Laboratory for Perceptual Robotics
Computer Science Department
University of Massachusetts Amherst
{chao,grupen}@cs.umass.edu

***Abstract*— Assisting humans in their daily lives requires robots to be proficient in manual tasks and effective in communicating states/intentions with human users. This paper advocates a learning approach for the development of communicative behavior in robots and favors a uniform means of learning communicative actions and manual skills in the same framework. In fact, this work argues for a critical relationship between the structure of motor skills and the structure required to communicate effectively. We show how to reuse manual behavior for conveying intentions to humans and to do so in the same grounded manner as the robot learns to interact with other objects in the environment. The learning framework and preliminary human-robot interaction experiments are presented, where a humanoid robot incrementally builds and refines communicative actions by discovering the utility of manipulation behavior in the presence of humans. The learning results from 18 subjects provide support for the hypothesized benefits of our approach that behavior reuse made learning from relatively few interactions possible and the robust manual behavioral basis kept the subjects interested. The approach presented in this paper compliments other efforts in the field as it grounds social behaviors, allowing them to be more adaptive to context changes or variations in human user preferences.**

## I. INTRODUCTION

In order for robots to assist humans in daily activities, both at home and at work, current state of the art approaches incorporate social behavior into robots. It has been shown that these systems engage human collaborators and improve the effectiveness of human-robot interfaces. Social behavior in robots is usually designed to mimic human behavior, for instance, gaze-capture, pointing, request for object, nodding or shaking head [1][2][3][4]. This paper focuses on the origins of basic social behaviors and attempts to understand how some of these gestures can naturally arise from interactions with humans in the environment, without explicit programming.

We advocate a learning approach to communicative behavior for a number of reasons: (1) given different physical appearances, morphologies and capabilities of robots, simple mimickry may not be the most effective way to communicate with the human because the relationship of the gesture to intention in both the expressive and receptive agent is lost. For example, different robots may need different gestures to convey the same intention; (2) learning can be used to augment existing behavior, regardless whether it was previously learned or arises from prior programming. It is difficult for designers to anticipate all possible contexts of an "open" real-world environment. A change in context may cause humans to react differently. It is important for the robot to possess the ability to identify contexts and learn to adapt to them; (3) different users may have different perceptions of different gestures. Therefore a learning approach can potentially enable the robot's communicative behavior to be adaptive to the preference of the user. Overall, a learning approach to building communicative behavior is complimentary to the other higher-level efforts in the field: it grounds social behaviors in robots, making them adaptive to context changes and variations of human preferences.

Fig. 1. Robot learning to gesture in the presence of a human

One of the obstacles for a learning approach is that currently most advanced machine learning algorithms are best suited for offline processing of large datasets or simulation runs that generally require tens of thousands of training episodes [5]. For the domain of human-robot interaction (HRI), this is particularly problematic since in order to acquire training data, a human needs to be present. Tens of thousands of training episodes is out of the question. For HRI, there has been a great deal of work devoted to reducing the training time in the domain of teaching by demonstration. However, they are mostly focused on optimizing low-level motion trajectories to achieve tasks such as performing a tennis forehand swing[6], batting a table-tennis ball, or catching table-tennis ball in a cup. Similar work has been done on teaching robots to produce gestures, but again they either treat gesture learning as a low-level motion trajectory problem [7] or a joint space motor control problem [8]. None of these approaches considered the interplay between the robot and human as part of the gesture learning process, how

environmental changes may affect the meaning of gestures, and how the robot may learn to adapt accordingly.

This paper presents such a learning framework and attempts to address some of the afore mentioned issues including the origin, adaptivity, and learning efficiency in the development of communicative behavior for robots. The proposed approach draws inspirations from the psychology literature and studies human-robot interaction in the same framework designed to acquire manual skills. Furthermore, gesture learning can directly benefit from the ideas of developmental staging, hierarchical learning and skill generalization that already exist in the control literature [9]. More importantly, communicative actions can now reuse behavior that was learned in the context of manipulation and extend that behavior to suit the desire to communicate intentions.

The rest of the paper is organized as follows: first, we discuss views from the psychology literature regarding the emergence of communicative behavior and the development of manual behavior. Next, we focus on the issue of how we can apply this insight to the domain of human-robot interaction and present a unified framework for the acquisition of both manipulation skills and social interaction behavior. In Section IV, the design of learning experiments with human subjects as well as findings from our preliminary experiments are presented. Finally, in Section VI, the implications and potential benefits of the proposed approach are further discussed.

## II. Psychology Literature on Manipulation and Communication

Psychologists acknowledge a tight connection between communicative geseture and manual behavior. In the 1930s, Vygotsky noted that "...initially, pointing is nothing more than an unsuccessful attempt to grasp something..." [10]. In this case, a manipulation behavior is described as the origin of the communicative pointing action. As infants attempt to reach for out-of-reach objects, even though they inevitably fail, in the presence of a caregiver, the action is recognized and interpreted as the "intention" to acquire the object and thus the action becomes a gesture. When infants become older, more sophisticated abstract gestural actions begin to emerge as infant's manipulation skills continue to improve. For instance, it is common for infants to pretend to drink from an empty cup to indicate the desire for a drink. This later often evolves to pantomiming without a cup as the infant's understanding of semantic meanings of actions improve [11].

Greenfield [12] hypothesized links between the origins of tool use and language, and also suggested that manipulation behavior for tool use may have played a causal role in the evolution of gestural communication. Studies of the brain functions [13] through observations of apraxia patients [14][15] and more recently fMRI machines [16] all provide positive evidence for this theory. Furthermore, results from Gibson's study [15] suggest that the human infants' capacity to learn complex sequence of actions in manipulation tasks and subsequent interest in object-object relationships allowed humans to develop complex systems of communication, including language, since sequencing behavior (utterances) and associating the causal outcome are also the key to developing effective communication skills.

For this work, we apply this insight to the field of robotics to show that this general principle can be integrated into a general-purpose computational framework for enabling robots to learn gestures in an grounded manner. Importantly, we contend that these forms of communicative actions can be built into social behavior without first constructing a mental model of the human subject—it relies only on discovering the causal relationships between "gesturer" and "gesturee." The "gesture" begins as a motor-artifact, is recognized as a reliable means of causation, and ultimately is acknowledged as an effective means of communicating ones intentions—and is initiated and perhaps stylized to that purpose.

## III. The Learning Framework

### A. The Formation of Stable Human-Robot Dyads

To verify and exploit the hypothesis in well-controlled robot learning experiments, a learning framework that supports stable dyadic relationships between a human subject and a robotic learning agent is needed. Specifically, these conditions are *underactuation* and *mutual reward*. Underactuation specifies that there exist conditions when some of the agents in a human-robot team cannot independently achieve the goal—objects can be too heavy, for instance, for any agent to lift alone, or objects can be unreachable by some agents and reachable by others. Mutual reward conditions require that each agent in a human-robot team be rewarded for participating constructively in a dyadic relationship. The rewards can be different events for different agents, but the polarity of reward/penalty must be the same for motivated engagement to take place. For example, in the case where the object is too heavy for either the robot or the human to lift alone, when the robot conveys the intention of wishing to lift the object to the human who chooses to help, the robot is rewarded for lifting the object, while the benevolent human is rewarded for successfully helping the robot to achieve its goal. Next, we present the learning framework employed by this work—the control basis framework—and show it can be used for the formation of stable human-robot dyads.

### B. The Control Basis Framework

The control basis framework is a principled approach for robots to learn hierarchical behavioral programs given available sensory and motor resources. Using this framework, a designer can guide a robot's learning process by simply controlling the resources and external stimuli made available to the robot at different times, thus creating a series of increasingly challenging stages. The robot learns simple programs first and then later moves onto more challenging scenarios with the availability of programs learned in the previous stages. In Section IV, examples are given to demonstrate how this strategy allows the designer to extend experiments for teaching robot manual skills and create

conditions that lead to the emergence of communicative gestures.

### C. Control Actions and State Estimation

Primitive actions in the control basis framework are closed-loop feedback controllers constructed by combining a potential function $\phi \in \Omega_\phi$, with a feedback signal $\sigma \in \Omega_\sigma$, and motor variables $\tau \in \Omega_\tau$ into a control action $c(\phi, \sigma, \tau)$. The potential function $\phi(\sigma)$ is a scalar function (e.g., a *navigation* function) defined to satisfy properties that guarantee asymptotic stability. Multi-objective control actions are achieved in the control basis by combining control primitives using nullspace composition.

The dynamics $(\phi, \dot{\phi})$ created when a controller interacts with the task domain supports a natural discrete abstraction of the underlying continuous state space [17]. One simple discrete state definition based on *quiescence events* and controller relevance was proposed in [9]. Quiescence events occur when a controller reaches an attractor state in its potential. We define a predicate predicate $p(\phi, \dot{\phi})$ associated with controller $c(\phi, \sigma, \tau)$, whose possible values are: $\{X, -, 0, 1\}$. The "$-$" condition means that no target stimuli is present in the feedback signal, $\sigma$, and the environment does not afford that control action at that time. The unknown "$X$" condition occurs when a controller is not running and has no dynamics. The "0" occurs during the transient response of $c_i$ as it descends the gradient of its potential, and "1" represents quiescence. Given a collection of $n$ distinct primitive control actions, a discrete state-space $\mathcal{S} \equiv (p_1 \cdots p_n)$ can be automatically formulated.

### D. Complex Behavior through Hierarchical Learning

To drive the learning process, this framework defines a simple intrinsic reward function $\mathcal{R}$ where the agent receives a unit of reward when a controller state transitions from 0 to 1. Given the state and action spaces $\mathcal{S}$ and $\mathcal{A}$ defined by the set $\{\Omega_\phi, \Omega_\sigma, \Omega_\tau\}$ and the reward function $\mathcal{R}$, together they form a graphical model of behavior and control. Formulating the learning problem as a Markov Decision Process (MDP) allows a learning agent to estimate the value, $\Phi(s, a)$, of taking action $a$ in state $s$ using reinforcement learning (RL) [5]. Representing behavior in terms of a value function provides a natural hierarchical representation for control basis programs where attractor states of the value function, $\Phi$, capture quiescence events in the policy. As a result, the state of a program can be captured using the same state-predicate representation as above, even though that program may have its own complex transition dynamics.

With this hierarchical learning framework, it has been demonstrated that a humanoid robot can learn complex behavior incrementally, from SEARCHTRACK, to REACH-GRAB, to VISUALINSPECT through stages of development [9], where each behavior builds ontop of the behavior learned in the past stage.

### E. Adaptation through Generalization and Prospective Repair

The increasingly complex set of behaviors are learned under constrained contexts, i.e., the robot is only allowed to explore using one of its arms and the same object is always placed in the same location on the table. Once the robot has learned the behavior, it is presented with more challenging scenarios where objects are placed at different regions of the workspace, and various scales of objects are also used.

To adapt to new contexts, robots running the control basis rely on two techniques: (1) re-parameterization of existing sensorimotor resources, e.g. if reaching with one arm fails, try using the other arm, or somtimes even attempt with both arms; (2) identifying the hidden state information that causes the failure of the existing policy and then attempting to learn a new sequence of actions to amend the condition that leads to failure and thus the existing strategy can be reused again to achieve the goal.

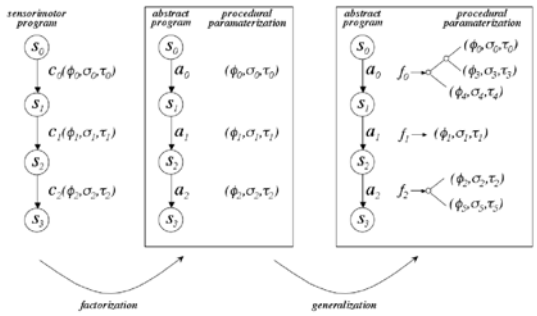


Fig. 2. Sensorimotor programs can be factored into abstract programs and procedural parameterizations such that the structure of the learned program can be generalized to new environmental contexts denoted as $f_i \in \mathcal{F}$ with only small amount of addition training.

The first technique has been demonstrated in an earlier paper by Hart [18]. The key of this approach is through factorization of the learned control program $\mathcal{C}$ into *declarative* components $a$ and *procedural* component $(\sigma, \tau)$ (Figure 2), where $a \in$ available action set $\mathcal{A}$, and $(\sigma, \tau)$ are sensorimotor resource pair (e.g. using the right arm, $\tau$, to reach to reference position, $\sigma$). This factorization enables the robot to quickly generalize to new contexts identified by the observed features $f$ and learn a mapping from $f$ to the appropriate sensorimotor resource for a given context, e.g. using left arm when object position feature $x_{obj}$ indicates the object on the left side ($x_{obj} < 0$) of the table.

There exists many situations where simple re-parameterization of existing behavior is not sufficient. In some cases, both the declarative structure and procedural knowledge need to be extended simultaneously for the behavior to be "repaired." In a recent paper [19], the control basis is extended with the prospective learning algorithm to handle such situations. The outline of the algorithm is summarized in Figure 3.

Through success and failure experiences gathered by the robot, the algorithm first identifies the environmental context

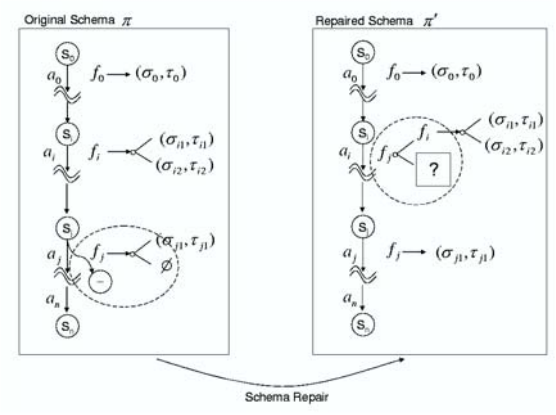


Fig. 3. Left: a context change $f_j$ alters the transitions of the existing policy $\pi$ that results in an unrewarding absorbing state '$-$' (dotted circle region on the left). Right: the prospective learning algorithm attempts to handle this context change by searching for fixes earlier on in the policy.

variable $f_j$ that causes the current behavior to fail (resulting in the non-rewarding absorbing '$-$' state) and learns a decision boundary $g$ for separating the two cases, i.e., $g = 0$ if the context predicts the behavior to fail and $g = 1$ where behavior will succeed. This allows a sub-learning problem to be automatically generated using conditions where $g : 0 \rightarrow 1$ as the goal. Then prospective learning back-tracks along the original policy until the earliest instance of the context of $f_j$ can be observed again and the robot explores its available actions and attempts to find action sequences that leads to the goal state ($g : 0 \rightarrow 1$), in a separate MDP generated from the available actions. After learning, the newly acquired sub-policy is merged with the original policy (Figure 3). Thus, prospective learning enables the robot adapt the new context while maintaining the structure of the previously learned program. A specific example of the algorithm in the context of learning communicative actions is given in Section IV. For further details and performance analysis of the algorithm, see [19].

## IV. Experimental Setup

To verify the proposed approach for robots to acquire communicative gestures, we employ a bimanual upper-torso humanoid robot, Dexter, as shown in Figure 1. Dexter has two 7-DOF Whole-Arm Manipulators (WAMs) and two 3-finger 4-DOF hands manufactured by Barrett Technologies. Each finger of the hand is equipped with one 6-axis force/torque load-cell sensor, enabling the robot to execute intricate grasping behavior. Other than forces, Dexter can also sense the world through a stereo camera pair mounted on a pan/tilt head.

For this work, the same experimental setup for learning manipulation skills is used and we simply extend the challenge by moving the objects further away until all objects are out-of-reach. An external resource, a benevolent human is introduced into the scene. This scenario naturally satisfies the conditions of *underactuation* and *mutual reward*. First, the robot is *underactuated* since it is unable to reach the desired object through its previously learned programs. However, when considering the human as part of the system, the system is *underactuated* because it is possible for the robot to influence the human to bring the object closer through some yet unknown sequence of actions. Secondly, the robot and the human are *mutually rewarded* since the robot is rewarded for touching the object, and on the other hand, the benevolent human is also rewarded when the robot touches the object since he/she finds it satisfying to assist the robot to achieve its goal, as long as the robot can correctly convey its intentions.

To faciliate the learning process, a previously used staged learning strategy is again applied by first limiting the robot's resources such that it is only allowed to explore actions associated with its head. In the second stage, this constraint is lifted and the robot is allowed to use both its arms and its head to explore. The goal of the experiment is to see if the learning framework enables the robot to learn sequences of actions that are useful for soliciting assistance from the human, even through these actions originally arose as the result of motor skill learning.

18 subjects of convenience are recruited for this study. Among these subjects, 7 are computer science students, including 2 lab members with extensive knowledge regarding the inner workings of Dexter. The remaining 11 are diverse in educational backgrounds in majors as well as level of education, ranging from high school students to undergrads, to graduate students and working professionals. The subjects are simply told to interact with robot for a number of rounds, and that "the robot will randomly pick an object of interest in each round, observe and help when necessary." All interactions between the robot and the subjects are recorded with consent for the purpose of offline analysis.

### A. Human Detection

In general, humans in the control basis learning framework are modeled the same way as the graspable objects on the table: the robot perceives visual features in space, and through the control basis it explores and finds actions associated with these features that lead to reliable rewarding controller transitions ($0 \rightarrow 1$). To facilitate Dexter's learning of humans in the environment, we biased the robot's visual sensory channels to be only sensitive at first to certain types of feature, e.g. large motions in the environment. This is similar to the maturational process of a human infant where at first the infant's vision is only responsive to large motions and brightly colored or high-contrast objects.

Using the control basis, Dexter autonomously creates controllers to explore these motion features associated with humans as it did before with the various objects on the table, and discovers several distinctive procedural properties related to humans: a) their color distributions change from day to day, i.e., the distribution has a large variance; b) their 3D positions over time forms a distinctive distribution that are different from the objects on the table; c) they are never graspable; d) their scales are different to the objects on the table. For the purpose of this work, these

distinctions are sufficient for both detection of humans in the environment, and learning basic communicative actions. For future work, acquiring finer kinematic model of humans for recognizing gestures from the human is also possible within this framework as we make finer features available to the robot. This discussion is beyond of the scope of this paper.

### B. Prospective Learning

At first, objects are placed out of reach of the robot without any humans in the environment. This gives Dexter the opportunity to learn about the length of its arm. The top of Figure 4 shows a simplified version of the REACHGRAB behavior program to illustrate the prospective learning process. As before, Dexter begins exploring local adjustments using a different resources, e.g. a different arm or both arms. When all valid resource sets are exhausted, the robot enters an unrewarding absorbing state $'-'$.

Statistics can be gathered on both successes and failures such that the contextual feature $f_j$ and a decision boundary $g$ that predicts failure can be identified using a standard discriminative learning algorithm (decision tree $C4.5$ is used here). In this case, $f_j$ corresponds to the $X$-axis in the robot's world coordinate frame, and the decision boundary $g : \{(X > 1.2) \rightarrow 0, (X <= 1.2) \rightarrow 1\}$ represents the procedural knowledge that failure '0' occurs when $X > 1.2$ meters. Through this process, the hidden state that causes the failure of the existing policy is uncovered. Thus, the transition s.t. $g : 0 \rightarrow 1$ becomes a new sub-goal for Dexter as $g = 1$ predicts success for the existing policy. At shown in the bottom of Figure 4, first, back-tracking is initiated on the existing policy $\pi$ to find the earliest state where $f_j$ can be observed. Next, a search for actions capable of causing the transition $g : 0 \rightarrow 1$ begins.

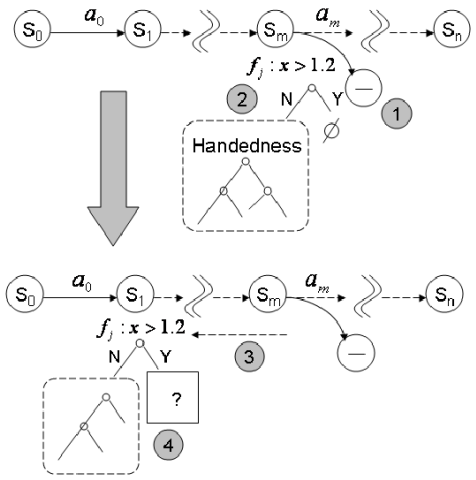


Fig. 4. When the object is out-of-reach, simply selecting a different arm is no longer sufficient. Thus the generalization procedure ensues: (1) the robot detects the failure as it enters an unrewarding absorbing $'-'$ state. (2) Through gathered experience, the robot then uncovers that when the object is at least some distance away ($x > 1.2m$), it can no longer reach the object. (3) The robot back-tracks in its learned REACHGRAB program and finds the earliest state where the context $x > 1.2m$ can also be observed, and (4) begins the search process for an action or actions such that allows the original program to continue and eventually succeed.

## V. RESULTS AND DISCUSSIONS

### A. Learning to use Gaze

As mentioned before, to faciliate learning, previously used strategy of developmental learning is applied. In this stage, Dexter is limited to explore actions using its head degrees of freedom and learn policies that cause the human to help. The only program Dexter has learned so far with its head is SACCADETRACK, denoted as $ST$ for short. However, Dexter has the option to proceduralize the program to attend to any objects in the environment: $ST(o_i)$, where $o_i \in O$, and set $O$ contains features that are associated with objects observed by Dexter.

As a human is enters the scene, Dexter becomes attentive to the large motion feature. It then tries to proceduralize its abstract behavior program to create actions that attends to the new feature and see if any action sequence can cause the desired object to appear closer. Therefore, for this stage, the actions made available to Dexter are a SACCADETRACK action attentive to the large motion feature $\{ST(human)\}$ and a similar action directed towards the desired object $\{ST(obj)\}$. However, since the sub-goal created by the prospective learning algorithm is related to the position of object, a monitor is also needed. In the control basis, a *monitor* is similarly configured as a controller with the exception that no effector resource is attached—it simply passively observes through its configured sensor resources. In this case, the monitor for the object is defined as $\phi_m^{obj_{cart}}$, where the lower script $m$ indicates this is a monitor, and $obj_{cart}$ is the triangulated position of the desired object in Cartesian space and the dynamic state of the monitor is '1' when the position of the object crosses the decision boundary $g : X < 1.2$ and '0' vice versa. For short, this monitor is denoted as $\phi_m^{obj}$. Therefore, the resulting action set $\mathcal{A}$ available to Dexter is: $\mathcal{A} \in \{ST(human), ST(obj) \triangleleft \phi_m^{obj}\}$, where the monitor is concurrently executed with the SACCADETRACK action associated with the object. According to the control basis (Section III), from the action set $\mathcal{A}$, a 3-predicate state space $\mathcal{S}$ is automatically formed: $\mathcal{S} : \{p_{ST_{human}}, p_{ST_{obj}}, p_{m_{obj}}\}$, one predicate for each of the actions and monitor.

Given the state-action space, a goal and 1 unit of reward for achieving the goal, Dexter can explore and see if learning an action sequence causes the object to move closer, using standard Q-learning. There are number of different policies that Dexter can potentially learn: e.g. alternating gazes between the human and the object, look at the human then keep staring at the object, or simply keep staring at the object. Before the experiment, we expected any of them would be sufficient as a policy for soliciting the appropriate response from the human subject. However, after interactions with 5 different subjects, Dexter settled on a single policy that uses alternating gazes a means of acquiring assistance from the human subjects (Figure 5). The learning curve is shown in

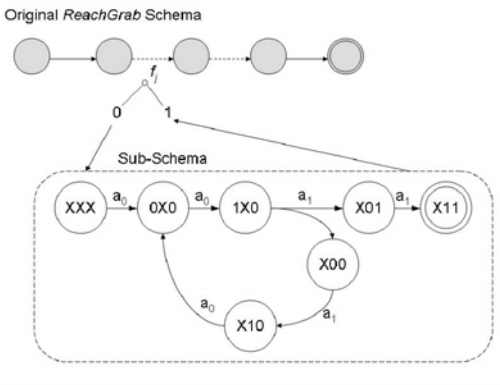


Fig. 5. New policy for REACHGRAB, with a "repair" sequence that resembles a gaze gesture acquired through prospective learning. In the repair policy MDP, $a_0$ corresponds to $ST_{human}$ and $a_1$ for $ST_{obj}$. Each state predicate in the MDP corresponds to the dynamic state of the action and monitor. The policy alternates SACCADETRACK actions directed at the human and the object in a cycle.

Figure 6.

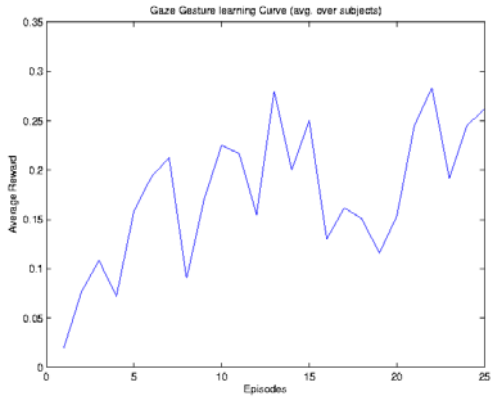


Fig. 6. Gaze gesture learning curve, averaged reward per state transition over all subjects. The first 15 episodes are the training phase while the remaining 10 episodes belong to the testing phase. The dip in average reward at the beginning of the testing phase is caused by the ambiguity of the gaze gesture where many subjects are initially confused about where to place the object.

Revisits of the recorded video footages reveal that it is reasonable that the learned policy won over the other two candidates. From the few attempts of the other two policies, i.e. look at the human and then keep staring at the object, or simply just keep staring at the object, did not cause the human to respond and therefore were not rewarded while the alternating gaze actions provoked some response from all subjects. This is because gazes are subtle movements that are often neglected if only executed once, and subjects did not realize the robot has performed any action and simply kept waiting for the robot to do something. On the other hand, alternating gazes are much more conspicuous and therefore led to more successes and quickly became favored as the greedy policy.

The learned policy is then tested on 10 subjects. Figure 7 shows that the new policy, with a "repair" sequence that corresponds to the gaze gesture. Even though the new policy is not yet ideal for acquiring the appropriate response from the human, it performed much better than the original policy that does nothing and the human subject would have to pick an object at random. Offline analysis of the recorded videos reveals interesting insights regarding the causes of the failed attempts: (1) gaze is imprecise as deictic pointers because the motion is subtle and therefore sometimes caused the human to pick the wrong/adjacent object. (2) Even when the subject has determined correctly which object the robot wants, it can still be ambiguous as to where the object needs to be placed. $60\%$ of the subjects took several tries to place the object within the reachable region of the robot. For these reasons, one subject showed confusion about the gaze gesture throughout of his interaction with Dexter, and managed to help only once.

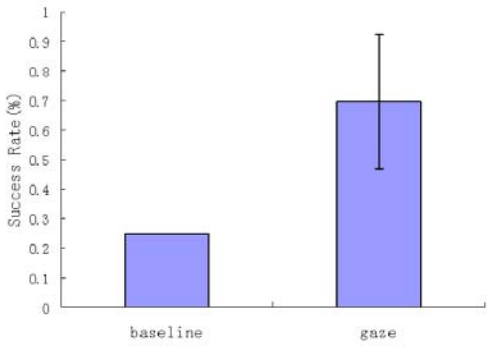


Fig. 7. Learned gaze gesture performance for acquiring human assistance compared against the baseline where the human can only pick an object at random to hand to the robot. The expected success rate ($0.25$ for $4$ objects) is used as the baseline.

Even more surprisingly, for this stage, novice subjects had more successful rounds of interaction than supposedly more "knowledgeable" students with robot experience (Figure 8). A possible explanation is that since this is a such a simple scenario, over-analyzing (speculating on how Dexter receives reward, or what actions Dexter will take) tends to cause more confusion and hesitation than if the subject simply acted out instinctively. The result is even more significant when we further divide the "knowledgeable" students into two categories, one group contains subjects who have worked with Dexter and the other contains the rest. The "Dexter-experienced" group performed worse than the other because they are used to Dexter gazing at objects with one of its eyes and therefore attempted to parse the direction of the gaze using the dominant eye. However, unknown to them, for this experiment, Dexter was configured to track using both of its cameras and as a result, its gaze direction keeps the object in-between its eyes. One of these "Dexter-experienced" subjects realized this in the middle of the experiment and corrected accordingly, while the other persisted till the end and made quite a few wrong guesses, thus lowering the overall statistics.

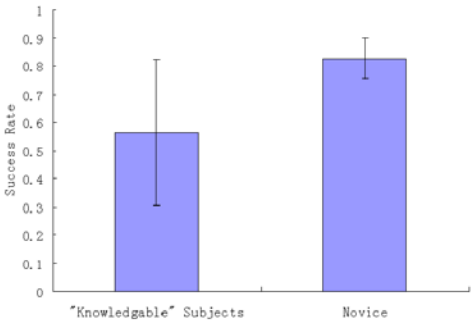


Fig. 8. Comparison between "Knowledgable" subjects with robot experience and naive subjects

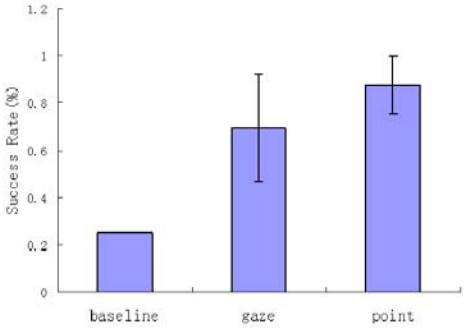


Fig. 10. Pointing policy performance in comparison with the previously learned gaze policy

### B. Learning Arm Pointing

During the second stage, Dexter was allowed to explore using its arms and previous manipulation behavior through out its interaction with the subjects. Therefore, the resulting action set $\mathcal{A}_2$ available to Dexter is: $\mathcal{A}_2 \in \{ST(human),\ ST(obj),\ RG(obj) \triangleleft \phi_m^{obj}\}$, where $RG$ denotes the learned REACHGRAB manipulation program that contains its own internal MDP. The state space is therefore: $\mathcal{S} : \{p_{ST_{human}}, p_{ST_{obj}}, p_{RG_{obj}}, p_{m_{obj}}\}$, where $m_{obj}$ is the monitor predicate.

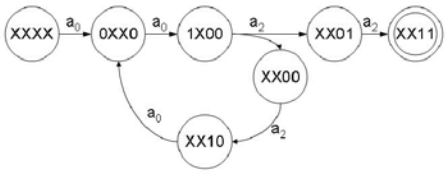


Fig. 9. Pointing gesture policy for repairing the original REACHGRAB program. The robot has learned to alternate between gazing at the human ($a_0$ is $ST(human)$) and reaching for the object ($a_2$ corresponds to the REACHGRAB action at the object). Each state predicate in the MDP corresponds to the dynamic state of the actions and monitor in $\mathcal{A}_2$.

Due to the introduction of extra resources—the arms, the number of actions available to the robot has increased and thus the state-action space grows. As expected, the training sessions took longer. However, Dexter still learned a useful policy (Fig. 9) within a reasonable 30 training episodes with 3 subjects. This is because due to developmental structuring, the problem is simple, the resulting policy used only two actions and has a simplistic structure and therefore is easy stumble upon.

Interestingly, the resulting policy has the same structure as the previously learned gaze gesture. This implies that if we reuse the structure of gaze gesture and simply swap out the $ST(obj)$ with $RG(obj)$ after learning the gaze gesture, it is possible to obtain a skeleton of the arm pointing gesture with no additional training. Of course, further training can be performed and it may be refined over time.

For this experiment, training was carried out with 3 subjects while the learned policy was tested on 8 subjects. 3 subjects in this experiment overlapped with the previous gaze experiment. This time, as a natural outcome of exploring learned manipulation behavior, Dexter found the failed attempt to reach and grab the desired object a more effective alternative for the gaze gesture (Figure 10). This is expected because the arm pointing gesture is both more conspicuous for capturing the subject's intention, and causes less ambiguity regarding which is the desired object and where it needs to be placed. In fact, even the subject who failed to attend to Dexter in the previous stage, responded almost immediately in this stage.

### C. Potential Issues of the Pointing Gesture

An unexpected result from the pointing experiment was that it revealed a pathological flaw of the learned pointing gesture: when the human handed the object to the robot's out-stretched hand, sometimes the object can be blocked by the hand. As a result, the robot retracted its arm and thus confusing the subject, who thought he selected the wrong object. Although this did not occur often enough to prevent the robot from learning the pointing gesture, it is conceivable that if smaller objects are used, more unsuccessful attempts would arise.

This problem can be resolved if the robot develops the understanding of occlusion as part of its manipulation skill set. Or, it is also possible for the robot to keep exploring other manipulation behavior, and hopefully it can find an alternative policy without the pathological issues of the current pointing gesture. One such possible alternative could be achieved when a new manipulation behavior, i.e., PICKANDPLACE, becomes available through manual skill learning. This is because when parameterized properly, the *pick* goal of the PICKANDPLACE action can indicate the object of desire, while the *place* goal designates the placement location. Thus reducing the likelihood of occlusion that exist for the pointing gesture. Such learning is possible within the proposed staged learning approach.

### D. Maintaining Human Interest

For this set of experiments we assume by default that the human subjects are benevolent and therefore should always

behave to help the robot whenever possible. We also made sure the training sessions are short enough that most human subjects do not lose patience and violate the benevolent human assumption.

During the course of the experiments, we noticed that for most subjects, once they discovered the general strategy for recognizing the robot's intention, they behaved deterministically and patiently repeated the strategy (e.g. keep placing the object in the same place) until all required rounds are completely. For these subjects, the general assumption of benevolent humans applies and thus the mutual reward condition is automatically met.

However, 2 subjects behaved differently. They soon exhibited signs of boredom after discovering general strategy for helping the robot, and started experimenting different options to test the capability of the robot by hiding the desired object from the robot's view, placing the object at random locations, moving the object while the robot is pointing, or attempting different intial object configurations by swapping objects around or stacking them up. Due to the robustness of existing behavioral programs, Dexter was able to handle most of testing situations posed by the human and acted "sensibly", i.e. using the left hand for objects placed on the left side and the right hand for objects placed on the right, and the "point" dynamically followed the object if it is moved. This intentional testing kept the subjects interested and even motivated one to perform 5 more rounds of training beyond the nominally required amount. These observations lend support for the use of existing manual behavior as the basis of communicative gestures as our results suggest that a robot with high level of aptitude in manual skills keeps the human mutually rewarded and thus allowing the human-robot dyad to be maintained.

## VI. Conclusion and Discussion

This paper presents a principled, grounded approach toward the acquisition of expressive communicative behavior for robots and presents a framework that enables robots to learn communicative actions and manual skills in conjunction. Human subject experiments demonstrate the feasibility of this approach where a humanoid robot built and refined its communicative behavior repertoire for acquiring human assistance, in a scenario where the desired object is placed out of reach of the robot. The approach enabled the reuse of manual skills acquired from previous sessions where it learns robust behavior for interacting with objects in the environment.

Using these manual behaviors as the basis of communicative gesture learning, this work demonstrated that with very few on-line interactions with the human subjects, the robot was able to learn behavior programs that effectively convey its intentions to humans. Through stages of learning, the robot also exhibited an incremental learning process in developing its gestural skill set, where it initially learned a somewhat ambiguous/less effective gaze gesture, then later developed the pointing gesture. Possible learning stages to further improve the effectiveness of the pointing gesture are also suggested. Furthermore, the experiments also provided positive evidence for using robust manipulation behavior as the basis of social interaction behavior can be beneficial for maintaining interest of the human and thus prolonging the interaction experience. These results both provide support for the approach to connect manual and communicative behavior learning, and offer interesting insights to the development of communicative behavior in robots, learned or otherwise.

## VII. Acknowledgments

This research is supported under the NASA-STTR-NNX08CD41P, ARO-W911NF-05-1-0396, and ONR-5710002229. The authors would also like to acknowledge Stephen Hart and Shiraj Sen for their helpful discussions.

## References

[1] C. Breazeal, A. Brooks, J. Gray, G. Hoffman, C. Kidd, J. Lieberman, A. Lockerd, and D. Mul, "Humanoid robots cooperative partners people," *International Journal Humanoid Robots*, 2004.

[2] B. Mutlu, "A storytelling robot: Modeling and evaluation of human-like gaze behavior. under review," in *International Conference on Humanoid Robots*, 2006.

[3] A. Edsinger and C. Kemp, "Human-robot interaction for cooperative manipulation: Handing objects to one another," in *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication (ROMAN)*, 2007.

[4] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior," in *HRI '09: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, 2009.

[5] R. Sutton and A. Barto, *Reinforcement Learning*. Cambridge, Massachusetts: MIT Press, 1998.

[6] C. Atkeson and S. Schaal, "Robot learning from demonstration," in *Intenational Conference on Machine Learning*, 1997.

[7] S. Calinon and A. Billard, "Stochastic gesture production and recognition model for a humanoid robot," in *Proceedings of the international Conference on Intelligent Robots and Systems*, 2004, pp. 2769–2774.

[8] G. S. M. Doniec and B. Scassellati, "Active learning of joint attention," in *IEEE/RSJ International Conference on Humanoid Robotics*, 2006.

[9] S. Hart, S. Sen, and R. Grupen, "Intrinsically motivated hierarchical manipulation," in *Proceedings of 2008 IEEE Conference on Robots and Automation (ICRA)*, 2008.

[10] L. Vygotsky, *Mind in society*. Harvard University Press, 1930.

[11] E. Bates and F. Dick, "Language, gesture, and the developing brain," *Developmental Psychobiology*, 2002.

[12] P. Greenfield, "Language, tools and the brain: the development and evolution of hierarchically organized sequential behavior." *Behavioral and Brain Sciences*, vol. 95, p. 531, 1991.

[13] D. Kimura, "Neuromotor mechanisms in the evolution of human communication," *In Neurobiology of Social Communication in Primates: An Evolutionary Perspective*, 1979.

[14] J. Bradshaw and L. Rogers, *The evolution of lateral asymmetries, language, tool use, and intellect*. Academic Press, 1993.

[15] K. Gibson and T. Ingold, *Tools, language and cognition in human evolution*. Cambridge University Press, 1993.

[16] S. H. Frey, "Tool use, communicative gesture and cerebral asymmetries in the modern human brain," *Philosophical Transactions of the Royal Society*, 2008.

[17] M. Huber, "A hybrid architecture for adaptive robot control," Ph.D. dissertation, Department of Computer Science, University of Massachusetts Amherst, 2000.

[18] S. Hart, S. Sen, and R. Grupen, "Generalization and transfer in robot control," in *Epigenetic Robotics Annual Conference*, 2008.

[19] S. Ou and R. Grupen, "Learning prospective robot behavior," in *AAAI Spring Symposium, Agents that learn from Human Teachers*, 2009.