Shared Intentional Plans for Imitation and Cooperation: Integrating Clues from Child Development and Neurophysiology into Robotics

Peter Ford Dominey

Abstract. One of the long-term goals in the domain of human-robot interaction is that robots will approach these interactions equipped with some of the same fundamental cognitive capabilities that humans use. This will include the ability to perceive and understand human action in terms of an ultimate goal, and more generally to represent shared intentional plans in which the goal directed actions of the robot and the human are interlaced into a shared representation of how to achieve a common goal in a cooperative manner. The current research takes specific experimental protocols from studies of cognitive development to define behavior milestones for a perceptual-motor robotic system. Based on a set of previously established principals for defining the "innate" functions available to such a system, a cognitive architecture is developed that allows the robot to perform cooperative tasks at the level of an 18 month old human child. Structural and functional properties of the primate neurophysiological mechanisms for action processing are used to provide further constraints on how the architecture is implemented. At the interface of cognitive development and robotics, the results on cooperation and imitation provide (1) a concrete demonstration of how cognitive neuroscience and developmental studies can contribute to human-robot interaction fidelity, and (2) a demonstration of how robots can be used to experiment with theories on the implementation of cognition in the developing human.

1. INTRODUCTION

One of the current open challenges in cognitive computational neuroscience is to understand the neural basis of the human ability to observe and imitate action. The results from such an endeavor can then be implemented and tested in robotic systems. Recent results from human and non-human primate behavior, neuroanatomy and neurophysiology provide a rich set of observations that allow us to constrain the problem of how imitation is achieved. The current research identifies and exploits constraints in these three domains in order to develop a system for goal directed action perception and imitation.

An impressive body of research exists on human imitation (62K responses to "human imitation" in Google Scholar), which has been empirically studied for over 100 years [15]. One of the recurrent findings across these studies is that in the context of goal directed action, it is the goal itself that tends to take precedence in defining what is to be imitated, rather than the means [1, 6, 825, 27, 28]. Of course in some situations it is the details (e.g. kinematics) of the movement itself that are to be imitated (see discussion in [6, 7]), but the current research focuses on goal based imitation . This body of research helped to formulate questions concerning what could be the neurophysiological substrates for goal based imitation. In 1992 di Pellegrino in the Rizzolatti lab [8] published the first results on "mirror" neurons, whose action potentials reflected both the production of specific goal-directed action, and the perception of the same action being carried by the experimenter. Since then, the premotor and parietal mirror system has been studied in detail in monkey (by single unit recording) and in man (by PET and fMRI) [see 25 for review].

In the context of understanding imitation, the discovery of the mirror system had an immense theoretical impact, as it provided justification for a common code for action production and perception. In recent years a significant research activity has used simulation and robotic platforms to attempt to link imitation behavior to the underlying neurophysiology at different levels of detail (see [24] for a recent and thorough review, edited volumes [22, 23], and a dedicated special issue of Neural Networks [2]). Such research must directly address the question of how to determine what to imitate. Carpenter and Call [6] distinguish three aspects of the demonstration to copy: the physical action, the resulting change in physical state, and the inferred goal – the internal representation of the desired state. Here we concentrate on imitation of the goal, with the advantage of eliminating the difficulties of mapping detailed movement trajectories across the actor and imitator [7].

Part of the novelty of the current research is that it will explore imitation in the context of cooperative activity in which two agents act in a form of turn-taking sequence, with the actions of each one folding into an interleaved and coordinated intentional action plan. With respect to constraints derived from behavioral studies, we choose to examine child development studies, because such studies provide well-specified protocols that test behavior that is both relatively simple, and pertinent. The expectation is that a system that can account for this behavior should extend readily to more complex behavior, as demonstrated below.

Looking to the developmental data, Warneken, Chen and Tomasello [30] engaged 18-24 month children and young chimpanzees in goal-oriented tasks and social games which required cooperation. They were interested both in how the cooperation would proceed under optimal conditions, but also how the children and chimps would respond when the adult had a problem in performing the task. The principal finding was that children enthusiastically participate both in goal directed cooperative tasks and social games, and spontaneously attempt to reengage and help the adult when he falters. In contrast, chimps are uninterested in non-goal directed social games, and appear

P. F. Dominey is with the CNRS, 67 Bd Pinel 69675 Bron Cedex, France (phone: 33-437-911266; fax: 33-437-9112110; e-mail: dominey@isc.cnrs.fr).

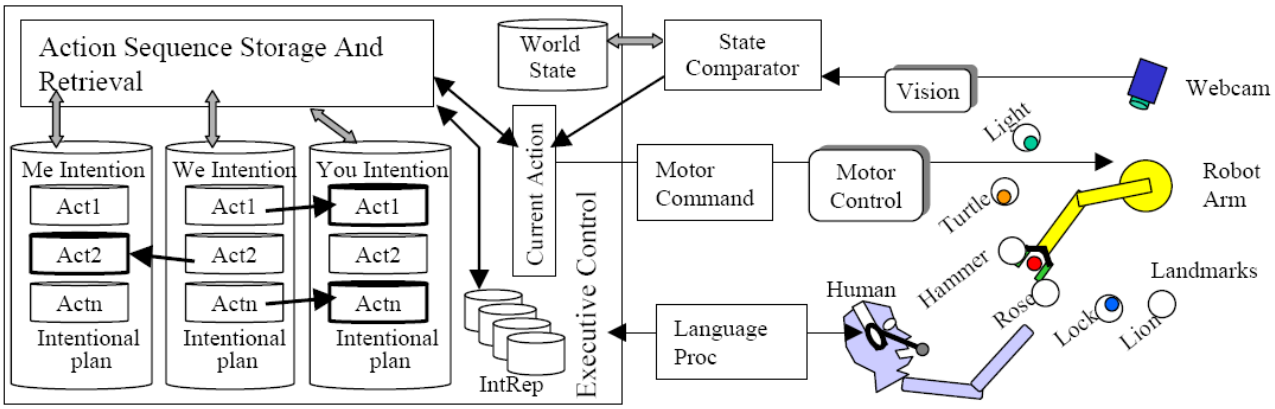

Fig 1. Cooperation System. In a shared work-space, human and robot manipulate objects (green, yellow, read and blue circles corresponding to dog, horse, pig and duck), placing them next to the fixed landmarks (light, turtle, hammer, etc.). *Action*: Spoken commands interpreted as individual words or grammatical constructions, and the command and possible arguments are extracted using grammatical constructions in Language Proc. The resulting Action(Agent, Object, Recipient) representation is the Current Action. This is converted into robot command primitives (Motor Command) and joint angles (Motor Control) for the robot. *Perception*: Vision provides object location input, allowing action to be perceived as changes in World State (State Comparator). Resulting Current Action used for action description, imitation, and cooperative action sequences. *Imitation*: The user performed action is perceived and encoded in Current Action, which is then used to control the robot under the supervision of Executive Control. *Cooperative Games*. During observations, individual actions are perceived, and attributed to the agent or the other player (Me or You). The action sequence is stored in the We Intention structure, that can then be used to separately represent self vs. other actions..

wholly fixed on attaining food goals, independent of cooperation. Warneken et al. thus observed what appears to be a very early human capacity for (1) actively engaging in cooperative activities for the sake of cooperation, and (2) for helping or reengaging the perturbed adult [29, 30].

In one of the social games, the experiment began with a demonstration where one participant sent a wooden block sliding down an inclined tube and the other participant caught the block in a tin cup that made a rattling sound. This can be considered more generally as a task in which one participant moves an object so that the second participant can then in turn manipulate the object. This represents a minimal case of a coordinated action sequence. After the demonstration, in Trials 1 and 2 the experimenter sent the block down one of the tubes three times, and then switched to the other, and the child was required to choose the same tube as the partner. In Trials 3 and 4 during the game, the experimenter interrupted the behavior for 15 seconds and then resumed.

Behaviorally, children successfully participated in the game in Trials 1 and 2. In the interruption Trials 3 and 4 they displayed two particularly interesting types of response that were (a) to attempt to perform the role of the experimenter themselves, and/or (b) to reengage the experimenter with a communicative act. This indicates that the children had a clear awareness both of their role and that of the adult in the shared coordinated activity. This research thus identifies a set of behavioral objectives for robot behavior in the perception and execution of cooperative intentional action. Such behavior could, however, be achieved in a number of possible architectures.

In order to begin to constrain the space of possible solutions we can look to recent results in human and primate neurophysiology and neuroanatomy. It has now become clearly established that neurons in the parietal cortex and the premotor cortex encode the goal of simple actions both for the execution of these actions as well as for the perception of these same goal-directed actions when performed by a second agent [8, 25]. This research thus corroborates the emphasis from behavioral studies on the importance of the goal (rather than the details of the means) in action perception [1, 6, 825, 27, 28]. It has been suggested that these "mirror" neurons play a crucial role in imitation, as they provide a common representation for the perception and subsequent execution of a given action. Interestingly, however, it has been clearly demonstrated that the imitation ability of non-human primates is severely impoverished when compared to that of humans [25, 28-30]. This indicates that the human ability to imitate novel actions and action sequences in real time (i.e. after only one or two demonstrations) relies on additional neural mechanisms.

In this context, a recent study of human imitation learning [5] implicates Brodmann's area (BA) 46 as responsible for orchestrating and selecting the appropriate actions in novel imitation tasks. We have recently proposed that BA 46 participates in a dorsal stream mechanism for the manipulation of variables in abstract sequences and language [14]. Thus, variable "slots" that can be instantiated by arbitrary motor primitives during the observation of new behavior sequences, are controlled in BA 46, and their sequential structure is under the control of corticostriatal systems which have been clearly implicated in sensorimotor sequencing (see [14]). This allows us to propose that this evolutionarily more recent cortical area BA 46 may play a crucial role in allowing humans to perform compositional operations (i.e. sequence learning) on more primitive action representations in the ventral premotor and parietal motor cortices. In other words, ventral premotor and parietal cortices instantiate shared perceptual and motor representations of atomic actions, and BA46 provides the capability to compose arbitrary sequences of these atomic actions, while relying on well known corticostriatal neurophysiology for sequence storage and retrieval. The functional result is the human ability to observe and represent novel behavioral action sequences. We further claim that this system can represent behavioral sequences from the "bird's eye view" or third person perspective, as required for the cooperative tasks of Warneken et al. [30]. That is, it can allow one observer to perceive and form an integrated representation of the coordinated actions of two other agents engaged in a cooperative activity. The observer can then use this representation to step in and play the role of either of the two agents.

2. IMPLEMENTATION

In a comment on Tomasello et al [28] on understanding and sharing intention, Dominey [10] analyses how a set of initial capabilities can be used to provide the basis for shared intentions. This includes capabilities to

1. perceive the physical states of objects,
2. perceive (and perform) actions that change these states,
3. distinguish between self and other,
4. perceive emotional/evaluation responses in others, and
5. learn sequences of predicate-argument representations.

The goal is to demonstrate how these 5 properties can be implemented within the constraints of the neurophysiology data reviewed above in order to provide the basis for performing these cooperative tasks. In the current experiments the human and robot cooperate by moving physical objects to different positions in a shared work-space as illustrated in Figures 1 and 2. The 4 moveable objects are pieces of a wooden puzzle, representing a dog, a pig, a duck and a cow. These pieces can be moved by the robot and the user in the context of cooperative activity. Each has fixed to it a vertically protruding metal screw, which provides an easy grasping target both for the robot and for humans. In addition there are 6 images that are fixed to the table and serve as landmarks for placing the moveable objects, and correspond to a light, a turtle, a hammer, a rose, a lock and a lion, as partially illustrated in Figures 1 & 2. In the interactions, human and robot are required to place objects in zones next to the different landmarks, so that the robot can more easily determine where objects are, and where to grasp them. Figure 1 provides an overview of the architecture, and Figure 2, which corresponds to Experiment 6 provides an overview of how the system operates.

2.1 Representation

The structure of the internal representations is a central factor determining how the system will function, and how it will generalize to new conditions. Based on the neurophysiology reviewed above, we use a common representation of action for both perception and production. Actions are identified by the agent, the object, and the target location to move that object to. As illustrated in Figure 1, by taking the short loop from vision, via Current Action Representation, to Motor Command, the system is thus configured for a form of goal-centered action imitation. This will be expanded upon below.

A central feature of the system is the World Model that represents the physical state of the world, and can be accessed and updated by vision, motor control, and language, similar to the Grounded Situation Model of [21]. The World Model encodes the physical locations of objects that is updated by vision and proprioception (i.e. robot action updates World Model with new object location). Changes in the World Model in terms of an object being moved allows the system to detect actions in terms these object movements. Actions are represented in terms of the agent, the object and the goal of the action, in the form MOVE(object, goal location, agent). These representations can be used for commanding action, for describing recognized action, and thus for action imitation and narration, as seen below.

In order to allow for more elaborate cooperative activity, the system must be able to store and retrieve actions in a sequential structure. This form of real time sequence learning for imitation is not observed in non-human primates. Interestingly, in this context, an fMRI study [5] that addressed the human ability to observe and program arbitrary actions indicated that a cortical area (BA46) which is of relatively recent phylogenetic origin is involved in such processes. Rizzolatti and Craighero [25] have thus suggested that the BA 46 in man will orchestrate allow the real-time capability to store and retrieve recognized actions, and we can further propose that this orchestration will recruit canonical brain circuitry for sequence processing including the cortico-striatal system (see [14] for discussion of such sequence processing).

In the current study we address behavioral conditions in which focus on the observation and immediate re-use of an intentional (goal directed) action plan. However, in the more general case, one should consider that multiple intentional action plans can be observed and stored in a repertory (IntRep or Intentional Plan Repertory in Figure 1). When the system is subsequently observing the behavior of others, it can compare the ongoing behavior to these stored sequences. Detection of a match with the beginning of a stored sequence can be used to retrieve the entire sequence. This can then be used to allow the system to "jump into" the scenario, to anticipate the other agent's actions, and/or to help that agent if there is a problem.

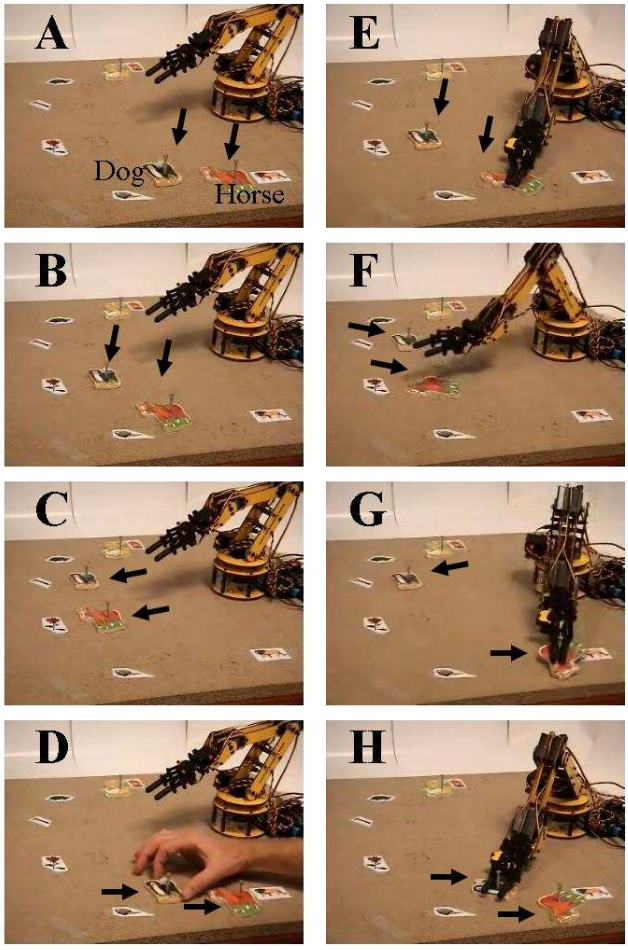

Figure 2. Cooperative task of Exp 5-6. Robot arm, with 6 landmarks (Light, turtle, hammer, rose, lock and lion from top to bottom). Moveable objects include Dog and Horse. In A-D, human demonstrates a "horse chase the dog" game, and successively moves the Dog then Horse, indicating that in the game, the user then the robot are agents, respectively. After demonstration, human and robot "play the game". In each of E – F user moves Dog, and robot follows with Horse. In G robot moves horse, then in H robot detects that the user is having trouble and so "helps" the user with the final move of the dog. See Exp 5 & 6.

2.2 Visual perception

Visual perception is a challenging technical problem. To simplify, standard lighting conditions and a small set (n = 10) of visual object to recognize are employed (4 moveable objects and 6 location landmarks). A VGA webcam is positioned at 1.25 meters above the robot workspace. Vision processing is provided by the Spikenet Vision System (http://www.spikenet-technology.com/). Three recognition models for each object at different orientations (see Fig. 3) were built with an offline model builder. During real-time vision processing, the models are recognized, and their (x, y) location in camera coordinates are provided. Our vision post-processing eliminates spurious detections and returns the reliable (x, y) coordinates of each moveable object in a file. The nearest landmark is then calculated.

2.3 Motor Control & Visual-Motor Coordination

While visual-motor coordination is not the focus of the current work, it was necessary to provide some primitive functions to allow goal directed action. All of the robot actions, whether generated in a context of imitation, spoken command or cooperative interaction will be of the form *move(x to y)* where *x* is a member of a set of visually perceivable objects, and *y* is a member of the set of fixed locations on the work plan.

Robot motor control for transport and object manipulation with a two finger gripper is provided by the 6DOF Lynx6 arm (www.lynxmotion.com). The 6 motors of the arm are coordinated by a parallel controller connected to a PC computer that provides transmission of robot commands over the RS232 serial port.

Human users (and the robot) are constrained when they move an object, to place it in one of the zones designated next to each of the six landmarks (see Fig 3). This way, when the nearest landmark for an object has been determined, this is sufficient for the robot to grasp that object at the prespecified zone.

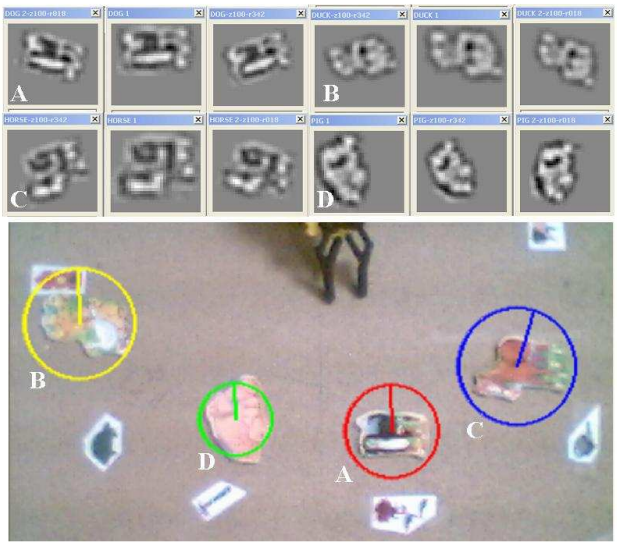

Figure 3. Vision processing. Above: A. – D. Three templates each for the Dog, Duck, Horse and Pig objects at three different orientations. Below, encompassing circles indicate template recognition for the four different objects near different fixed landmarks, as seen from the camera over the robot workspace

In a calibration phase, a target point is marked next to each of the 6 fixed landmark locations, such that they are all on an arc that is equidistant to the center of rotation of the robot arm base. For each, the rotation angle of Joint 0 (the rotating shoulder base) necessary to align the arm with that point is then determined, along with a common set of joint angles for Joints 1 – 5 that position the gripper to seize any of the objects. Angles for Joint 6 that controls the closing and opening of the gripper to grasp and release an object were then identified. Finally a neutral position to which the arm could be returned in between movements was defined. The system was thus equipped with a set of primitives that could be combined to position the robot at any of the 6 grasping locations, grasp the corresponding object, move to a new position, and place the object there.

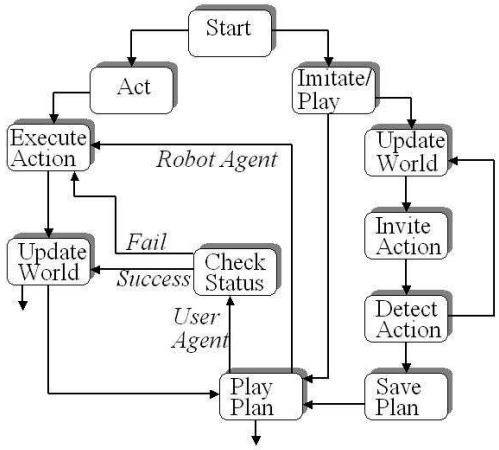

Figure 4. Spoken Language Based Cooperation flow of control. Interaction begins with proposal to act, or imitate/play a game. Act – user says an action that is verified and executed by robot. World Model updated based on action. Downward arrow indicates return to Start. Imitate/Play – user demonstrates actions to robot and says who the agent should be when the game is to be played (e.g. "You/I do this"). Each time, system checks the state of the world, invites the next action and detects the action based on visual object movement. When the demo is finished, the plan (of a single item in the case of imitation) is stored and executed (Play Plan). If the user is the agent (encoded as part of the game sequence), system checks execution status and helps user if failure. If robot is agent, system executes action, and then moves on to next item.

2.4 Cooperation Control Architecture

The spoken language control architecture illustrated in Fig 4 is implemented with the CSLU Rapid Application Development toolkit (http://cslu.cse.ogi.edu/toolkit/). This system provides a state-based dialog management system that allows interaction with the robot (via the serial port controller) and with the vision processing system (via file i/o). It also provides the spoken language interface that allows the user to determine what mode of operation he and the robot will work in, and to manage the interaction via spoken words and sentences.

Figure 4 illustrates the flow of control of the interaction management. In the Start state the system first visually observes where all of the objects are currently located. From the start state, the system allows the user to specify if he wants to ask the robot to perform actions (Act), to imitate the user, or to play (Imitate/Play). In the Act state, the user can specify actions of the form "Put the dog next to the rose" and a grammatical construction template [9,

11-14] is used to extract the action that the robot then performs. In the Imitate state, the robot first verifies the current state (Update World) and then invites the user to demonstrate an action (Invite Action). The user shows the robot one action. The robot re-observes the world and detects the action based on changes detected (Detect Action). This action is then saved and transmitted (via Play the Plan with Robot as Agent) to execution (Execute action). A predicate(argument) representation of the form Move(object, landmark) is used both for action observation and execution.

Imitation is thus a minimal case of Playing in which the "game" is a single action executed by the robot. In the more general case, the user can demonstrate multiple successive actions, and indicate the agent (by saying "You/I do this") for each action. The resulting intentional plan specifies what is to be done by whom. When the user specifies that the plan is finished, the system moves to the Save Plan, and then to the Play Plan states. For each action, the system recalls whether it is to be executed by the robot or the user. Robot execution takes the standard Execute Action pathway. User execution performs a check (based on user response) concerning whether the action was correctly performed or not. If the user action is not performed, then the robot communicates with the user, and performs the action itself. Thus, "helping" was implemented by combining an evaluation of the user action, with the existing capability to perform a stored action representation.

3. EXPERIMENTAL RESULTS

For each of the 6 following experiments, equivalent variants were repeated at least ten times to demonstrate the generalized capability and robustness of the system. In less than 5 percent of the trials, errors of two types were observed to occur. Speech errors resulted from a failure in the voice recognition, and were recovered from by the command validation check (Robot: "Did you say …?"). Visual image recognition errors occurred when the objects were rotated beyond 20° from their upright position. These errors were identified when the user detected that an object that should be seen was not reported as visible by the system, and were corrected by the user re-placing the object and asking the system to "look again". At the beginning of each trial the system first queries the vision system, and updates the World Model with the position of all visible objects. It then informs the user of the locations of the different objects, for example "The dog is next to the lock, the horse is next to the lion." It then asks the user "Do you want me to act, imitate, play or look again?", and the user responds with one of the action-related options, or with "look again if the scene is not described correctly.

3.1 Experiment 1: Validation of Sensorimotor Control

In this experiment, the user says that he wants the "Act" state (Fig 4), and then uses spoken commands such as "Put the horse next to the hammer". Recall that the horse is among the moveable objects, and hammer is among the fixed landmarks. The robot requests confirmation and then extracts the predicate-argument representation - *Move(X to Y)* - of the sentence based on grammatical construction templates. In the Execute Action state, the action *Move(X to Y)* is decomposed into two components of *Get(X)*, and *Place-At(Y)*. *Get(X)* queries the World Model in order to localize X with respect to the different landmarks, and then performs a grasp at the corresponding landmark target location. Likewise, *Place-At(Y)* simply performs a transport to target location Y and releases the object. Decomposing the *get* and *place* functions allows the composition of all possible combinations in the *Move(X to Y)* space. Ten trials were performed moving the four object to and from different landmark locations. Experiment 1 thus demonstrates (1) the ability to transform a spoken sentence into a Move(X to Y) command, (2) the ability to perform visual localization of the target object, and (3) the sensory-motor ability to grasp the object and put it at the specified location. In ten experimental runs, the system performed correctly.

3.2 Experiment 2: Imitation

In this experiment the user chooses the "imitate" state. As stated above, imitation is centered on the achieved ends – in terms of observed changes in state – rather than the means towards these ends. Before the user performs the demonstration of the action to be imitated, the robot queries the vision system, and updates the World Model (Update World in Fig 4) and then invites the user to demonstrate an action. The robot pauses, and then again queries the vision system and continues to query until it detects a difference between the currently perceived world state and the previously stored World Model (in State Comparator of Fig 1, and Detect Action in Fig 4), corresponding to an object displacement. Extracting the identity of the displaced object, and its new location (with respect to the nearest landmark) allows the formation of an *Move(object, location)* action representation. Before imitating, the robot operates on this representation with a meaning-to-sentence construction in order to verify the action to the user, as in "Did you put the dog next to the rose?" It then asks the user to put things back as they were so that it can perform the imitation. At this point, the action is executed (Execute Action in Fig 4). In ten experimental runs the system performed correctly. This demonstrates (1) the ability of the system to detect the goals of user-generated actions based on visually perceived state changes, and (2) the utility of a common representation of action for perception, description and execution.

3.3 Experiment 3: A Cooperative Game

The cooperative game is similar to imitation, except that there is a sequence of actions (rather than just one), and the actions can be effected by either the user or the robot in a cooperative manner. In this experiment, the user responds to the system request and enters the "play" state. In what corresponds to the demonstration in Warneken et al. [17] the robot invites the user to start showing how the game works. The user then begins to perform a sequence of actions. For each action, the user specifies who does the action, i.e. either "you do this" or "I do this". The intentional plan is thus stored as a sequence of action-agent pairs, where each action is the movement of an object to a particular target location. In Fig 1, the resulting interleaved sequence is stored as the "We intention", i.e. an action sequence in which there are different agents for different actions. When the user is finished he says "play the game". The robot then begins to execute the stored intentional plan. During the execution, the "We intention" is decomposed into the components for the robot (Me Intention) and the human (You intention).

In one run, during the demonstration, the user said "I do this" and moved the horse from the lock location to the rose location. He then said "you do this" and moved the horse back to the lock location. After each move, the robot asks "Another move, or shall we play the game?". When the user is finished demonstrating the

game, he replies “Play the game.” During the playing of this game, the robot announced “Now user puts the horse by the rose”. The user then performed this movement. The robot then asked the user “Is it OK?” to which the user replied “Yes”. The robot then announced “Now robot puts the horse by the lock” and performed the action. In two experimental runs of different demonstrations, and 5 runs each of the two demonstrated games, the system performed correctly. This demonstrates that the system can learn a simple intentional plan as a stored action sequence in which the human and the robot are agents in the respective actions.

3.4 Experiment 4: Interrupting a Cooperative Game

In this experiment, everything proceeds as in experiment 3, except that after one correct repetition of the game, in the next repetition, when the robot announced “Now user puts the horse by the rose” the user did nothing. The robot asked “Is it OK” and during a 15 second delay, the user replied “no”. The robot then said “Let me help you” and executed the move of the horse to the rose. Play then continued for the remaining move of the robot. This illustrates how the robot’s stored representation of the action that was to be performed by the user allowed the robot to “help” the user.

3.5 Experiment 5: A More Complex Game

Experiment 3 represented the simplest behavior that could qualify as a cooperative action sequence. In order to more explicitly test the intentional sequencing capability of the system, this experiment replicates Exp 3 but with a more complex task, illustrated in Figure 2. In this game (Table 1), the user starts by moving0 the dog, and after each move the robot “chases” the dog with the horse, till they both return to their starting places.

Action	User identifies agent	User Demonstrates Action	Ref in Figure 2
1.	I do this	Move dog from the lock to the rose	B
2.	You do this	Move the horse from the lion to the lock	B
3.	I do this	Move the dog from the rose to the hammer	C
4.	You do this	Move the horse from the lock to the rose	C
5.	You do this	Move the horse from the rose to the lion	D
6.	I do this	Move the dog from the hammer to the lock	D

Table 1. Cooperative “horse chase the dog” game specified by the user in terms of who does the action (indicated by saying) and what the action is (indicated by demonstration). Illustrated in Figure 2.

As in Experiment 3, the successive actions are visually recognized and stored in the shared “We Intention” representation. Once the user says “Play the game”, the final sequence is stored, and then during the execution, the shared sequence is decomposed into the robot and user components based on the agent associated with each action. When the user is the agent, the system invites the user to make the next move, and verifies (by asking) if the move was ok. When the system is the agent, the robot executes the movement. After each move the World Model is updated. As in Exp 3, two different complex games were learned, and each one “played” 5 times. This illustrates the learning by demonstration [31] of a complex intentional plan in which the human and the robot are agents in a coordinated and cooperative activity.

3.6 Experiment 6: Interrupting the Complex Game

As in Experiment 4, the objective was to verify that the robot would take over if the human had a problem. In the current experiment this capability is verified in a more complex setting. Thus, when the user is making the final movement of the dog back to the “lock” location, he fails to perform correctly, and indicates this to the robot. When the robot detects failure, it reengages the user with spoken language, and then offers to fill in for the user. This is illustrated in Figure 2H. This demonstrates the generalized ability to help that can occur whenever the robot detects the user is in trouble.

4. DISCUSSION

Significant progress has been made in identifying some of the fundamental characteristics of human cognition in the context of cooperative interaction, particularly with respect to social cognition [16-19]. Breazeal and Scassellati [4] investigate how perception of socially relevant face stimuli and object motion will both influence the emotional and attentional state of the system and thus the human-robot interaction. Scassellati [26] further investigates how developmental theories of human social cognition can be implemented in robots. In this context, Kozima and Yano [18] outline how a robot can attain intentionality – the linking of goal states with intentional actions to achieve those goals – based on innate capabilities including: sensory-motor function and a simple behavior repertoire, drives, an evaluation function, and a learning mechanism.

The abilities to observe an action, determine its goal and attribute this to another agent are all clearly important aspects of the human ability to cooperate with others. The current research demonstrates how these capabilities can contribute to the “social” behavior of learning to play a cooperative game, playing the game, and helping another player who has gotten stuck in the game, as displayed in 18-24 month children [29, 30]. While the primitive bases of such behavior is visible in chimps, its full expression is uniquely human [29, 30]. As such, it can be considered a crucial component of human-like behavior for robots.

The current research is part of an ongoing effort to understand aspects of human social cognition by bridging the gap between cognitive neuroscience, simulation and robotics [3, 9-14], with a focus on the role of language (see [20]). The experiments presented here indicate that functional requirements derived from human child behavior and neurophysiological constraints can be used to define a system that displays some interesting capabilities for cooperative behavior in the context of imitation. Likewise, they indicate that evaluation of another’s progress, combined with a representation of his/her failed goal provides the basis for the human characteristic of “helping.” This may be of interest to developmental scientists, and the potential collaboration between these two fields of cognitive robotics and human cognitive development is promising. The developmental cognition literature lays out a virtual roadmap for robot cognitive development [10, 28]. In this context, we are currently investigating the development of hierarchical means-end action sequences [27]. At each step, the objective will be to identify the behavior characteristic and to implement it in the most economic manner in this continuously developing system for human-robot cooperation.

At least two natural extensions to the current system can be considered. The first involves the possibility for changes in perspective. In the experiments of Warneken et al. the child watched two adults perform a coordinated task (one adult launching the block down the tube, and the other catching the block). At 24 months, the child can thus observe the two roles being played out, and then step into either role. This indicates a "bird's eye view" representation of the cooperation, in which rather than assigning "me" and "other" agent roles from the outset, the child represents the two distinct agents A and B for each action in the cooperative sequence. Then, once the perspective shift is established (by the adult taking one of the roles, or letting the child choose one) the roles A and B are assigned to me and you (or vice versa) as appropriate.

This actually represents a minimal change to our current system. First, rather than assigning the "you" "me" roles in the We Intention at the outset, these should be assigned as A and B. Then, once the decision is made as to the mapping of A and B onto robot and user, these agent values will then be assigned accordingly. Second, rather than having the user tell the robot "you do this" and "I do this" the vision system can be modified to recognize different agents who can be identified by saying their name as they act, or via visually identified cues on their acting hands.

The second issue has to do with inferring intentions. The current research addresses one cooperative activity at a time, but nothing prevents the system from storing multiple such intentional plans in a repertory (IntRep in Fig 1). In this case, as the user begins to perform a sequence of actions involving himself and the robot, the robot can compare this ongoing sequence to the initial subsequences of all stored sequences in the IntRep. In case of a match, the robot can retrieve the matching sequence, and infer that it is this that the user wants to perform. This can be confirmed with the user and thus provides the basis for a potentially useful form of learning for cooperative activity.

In conclusion, the current research has attempted to build and test a robotic system for interaction with humans, based on behavioral and neurophysiological requirements derived from the respective literatures. The interaction involves spoken language and the performance and observation of actions in the context of cooperative action. The experimental results demonstrate a rich set of capabilities for robot perception and subsequent use of cooperative action plans in the context of human-robot cooperation. This work thus extends the imitation paradigm into that of sequential behavior, in which the learned intentional action sequences are made up of interlaced action sequences performed in cooperative alternation by the human and robot. While many technical aspects of robotics (including visuomotor coordination and vision) have been simplified, it is hoped that the contribution to the study of imitation and cooperative activity is of some value.

5. ACKNOWLEDGEMENTS

I thank Mike Tomasello, Felix Warneken, Malinda Carpenter and Elena Lieven for useful discussions during a visit to the MPI EVA in Leipzig concerning shared intentions; and Giacomo Rizzolatti for insightful discussion concerning the neurophysiology of sequence imitation at the IEEE Humanoids meeting in Genoa 2006. This research is supported in part by the French Minister of Research under grant ACI-TTT, and by the LAFMI.

6. REFERENCES

[1] Bekkering H, WohlschlagerA , Gattis M (2000) Imitation of Gestures in Children is Goal-directed, The Quarterly Journal of Experimental Psychology: Section A, 53, 153-164

[2] Billard A, Schaal (2006) Special Issue: The Brain Mechanisms of Imitation Learning, Neural Networks, 19(1) 251-338

[3] Boucher J-D, Dominey PF (2006) Programming by Cooperation: Perceptual-Motor Sequence Learning via Human-Robot Interaction, Proc. Simulation of Adaptive Behavior, Rome 2006.

[4] Breazeal C., Scassellati B., (2001) Challenges in building robots that imitate people, in: K. Dautenhahn, C. Nehaniv (Eds.), Imitation in Animals and Artifacts, MIT Press, Cambridge, MA,.

[5] Buchine G, Vogt S, Ritzl A, Fink GR, Zilles K, Freund H-J, Rizzolatti G (2004) Neural circuits Underlying Imitation Learning of Hand Actions: An Event-Related fMRI Study. Neuron, (42) 323-334.

[6] Carpenter M, Call Josep (2007) The question of 'what to imitate': inferring goals and intentions from demonstrations, in Chrystopher L. Nehaniv and Kerstin Dautenhahn Eds, *Imitation and Social Learning in Robots, Human sand Animals*, Cambridge Univerity Press, Cambridge.

[7] Cuijpers RH, van Schie HT, Koppen M, Erlhagen W, Bekkering H (2006) Goals and means in action observation: A computational approach, *Neural Networks* 19, 311-322,

[8] di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G (1992) Understanding motor events: a neurophysiological study. *Exp Brain Res*.;91(1):176-80.

[9] Dominey, P.F., (2003) Learning grammatical constructions from narrated video events for human–robot interaction. *Proceedings IEEE Humanoid Robotics Conference*, Karlsruhe, Germany

[10] Dominey PF (2005) Toward a construction-based account of shared intentions in social cognition. Comment on Tomasello et al. 2005, *Beh Brain Sci*. 28:5, p. 696.

[11] Dominey PF, Alvarez M, Gao B, Jeambrun M, Weitzenfeld A, Medrano A (2005) Robot Command, Interrogation and Teaching via Social Interaction, *Proc. IEEE Conf. On Humanoid Robotics 2005*.

[12] Dominey PF, Boucher (2005) Learning To Talk About Events From Narrated Video in the Construction Grammar Framework, *Artificial Intelligence*, 167 (2005) 31–61

[13] Dominey, P. F., Boucher, J. D., & Inui, T. (2004). Building an adaptive spoken language interface for perceptually grounded human–robot interaction. In *Proceedings of the IEEE-RAS/RSJ international conference on humanoid robots*.

[14] Dominey PF, Hoen M, Inui T. (2006) A neurolinguistic model of grammatical construction processing. *Journal of Cognitive Neuroscience*.18(12):2088-107.

[15] Ellwood CA (1901) The Theory of Imitation in Social Psychology *The American Journal of Sociology*, Vol. 6, No. 6 (May, 1901), pp. 721-741

[16] Fong T, Nourbakhsh I, Dautenhaln K (2003) A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42 3-4, 143-166.

[17] Goga, I., Billard, A. (2005), Development of goal-directed imitation, object manipulation and language in humans and robots. In M. A. Arbib (ed.), *Action to Language via the Mirror Neuron System*, Cambridge University Press (in press).

[18] Kozima H., Yano H. (2001) A robot that learns to communicate with human caregivers, in: *Proceedings of the International Workshop on Epigenetic Robotics*,.

[19] Lieberman MD (2007) Social Cognitive neuroscience: A Review of Core Processes, *Annu. Rev. Psychol*. (58) 18.1-18.31

[20] Lauria S, Buggmann G, Kyriacou T, Klein E (2002) Mobile robot programming using natural language. *Robotics and Autonomous Systems* 38(3-4) 171-181

[21] Mavridis N, Roy D (2006). Grounded Situation Models for Robots: Where Words and Percepts Meet. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*

[22] Nehaniv CL, Dautenhahn K eds. (2002) Imitation in Animals and Artifacts; MIT Press, Cambridge MA.
[23] Nehaniv CL, Dautenhahn K eds. (2007) *Imitation and Social Learing in Robots, Humans and Animals*, Cambridge University Press, Cambridge.
[24] Oztop E, Kawato M, Arbib M (2006) Mirror neurons and imitation: A computationally guided review. *Neural Networks*, (19) 254-271
[25] Rizzolatti G, Craighero L (2004) The Mirror-Neuron system, *Annu. Rev. Neuroscience* (27) 169-192
[26] Scassellati B (2002) Theory of mind for a humanoid robot, *Autonomous Robots*, 12(1) 13-24
[27] Sommerville A, Woodward AL (2005) Pulling out the intentional structure of action: the relation between action processing and action production in infancy. *Cognition*, 95, 1-30.
[28] Tomasello M, Carpenter M, Cal J, Behne T, Moll HY (2005) Understanding and sharing intentions: The origins of cultural cognition, *Beh. Brain Sc*;. 28; 675-735.
[29] Warneken F, Tomasello M (2006) Altruistic helping in human infants and young chimpanzees, *Science*, 311, 1301-1303
[30] Warneken F, Chen F, Tomasello M (2006) Cooperative Activities in Young Children and Chimpanzees, *Child Development*, 77(3) 640-663.
[31] Zöllner R., Asfour T., Dillman R.: Programming by Demonstration: Dual-Arm Manipulation Tasks for Humanoid Robots. *Proc IEEE/RSJ Intern. Conf on Intelligent Robots and systems (IROS 2004).*