Cooperative Human Robot Interaction Systems: IV. Communication of Shared Plans with Naïve Humans using Gaze and Speech

Stéphane Lallée, Katharina Hamann, Jasmin Steinwender, Felix Warneken, Uriel Martienz, Hector Barron-Gonzales, Ugo Pattacini, Ilaria Gori, Maxime Petit, Giorgio Metta, Paul Verschure, Peter Ford Dominey

Abstract— Cooperation[1] is at the core of human social life. In this context, two major challenges face research on human-robot interaction: The first is to understand the underlying structure of cooperation, and the second is to build, based on this understanding, artificial agents that can successfully and safely interact with humans. Here we take a psychologically grounded and human-centered approach that addresses these two challenges. We test the hypothesis that optimal cooperation between a naïve human and a robot requires that the robot can acquire and execute a joint plan, and that it communicates this joint plan through ecologically valid modalities including spoken language, gesture and gaze. We developed a cognitive system that comprises the human-like control of social actions, the ability to acquire and express shared plans and a language and speech synthesis stage. This cognitive system was driving the actions of an iCub humanoid robot that maintained dyadic interactions with a single human actor. In order to test the psychological validity of our approach we tested 12 naïve subjects in a cooperative task with the robot. We experimentally manipulated the presence of a joint plan (vs. an individual or solo plan), the use of task-oriented gaze and gestures, and the use of language accompanying the unfolding plan. The quality of cooperation was analyzed in terms of proper turn taking, collisions and cognitive errors. Results showed that while successful turn taking could take place in the absence of the explicit use of a joint plan and/or its accompanying cues, the presence of a joint plan yielded significantly greater success. One advantage of the solo plan was that the robot would always be ready to generate actions, and could thus adapt if the human intervened at the wrong time, whereas in the joint plan the robot expected the human to take his/her turn. Interestingly, when the robot represented the action as involving a joint plan, gaze provided a highly potent nonverbal cue that facilitated successful collaboration and reduced errors in the absence of verbal communication. These results support the cooperative stance in human social cognition, and suggest that cooperative robots should employ joint plans, fully communicate them in order to sustain effective collaboration while being ready to adapt if the human makes a midstream mistake.

Stéphane Lallée and Paul Verschure are with SPECS, UPF & ICREA Barcelona, Spain, Katharina Hamann & Jasmin Steinwender are with the Max Planck Institute EVA Leipzig. Felix Warneken is with the Dept. Psychology, Harvard University. Uriel Martienz & Hector Barron-Gonzales are with the University of Sheffield, Ugo Pattacini, Illaria Gori & Giorgio Metta are with the Italian Institute of Technology, Maxime Petit & Peter Ford Dominey are with the Robot Cognition Laboratory, INSERM, Lyon. Peter.dominey@inserm.fr

Index Terms—cooperation, joint plan, shared intention, gaze, spoken language, HRI, cognitive architecture.

I. Introduction

A key challenge of robotics is to endow robots with the capability to collaborate closely with humans. This requires systems that can directly interact with humans while adapting to novel exigencies in the environment and responding to the inherently complex actions that humans perform. Despite these complexities, one biological system masters these challenges with apparent ease: human children. From early on in their lives, young children are able to socially interact with others in a cooperative fashion, demonstrating successful cooperation in fairly complex and sometimes novel situations – often without much learning and before they have developed a proper command of language or abstract thought [1, 2].

Research in human cognitive development has investigated the cognitive foundations at the basis of this capability to cooperate. Two aspects of human social cognition that stand out in this capability are (1) the capability to understand and represent others as intentional agents, and (2) the capability and motivation to share intentions. Together, these capabilities provide the basis for dialogic interactions centered on shared intentions, which lead to the construction of joint plans [3-5].

Joint plans correspond to representations created and negotiated by two agents that allow them to act together in a coordinated way to achieve their shared goal. Because of the supposed crucial role of joint plans in cooperative behavior, we have focused on the implementation of joint plan learning and use in the context of cooperative human-robot interaction [6-8]. Our previous research demonstrated that indeed, a robot equipped with the ability to learn and use joint plans could successfully learn new cooperative tasks and use the learned joint plan to perform the shared task with novel objects. In the current research, we extend this work in cooperation, and evaluate the psychological plausibility and efficiency of a human-like dyadic interaction based on joint plans expressed through gesture, gaze and speech. We test the hypothesis that optimal cooperation between a naïve human and robot requires that the robot has a joint plan, and that it communicates this joint plan through all modalities available including spoken language and gaze.

Dyadic social interaction is a central feature of human behavior that entails regular patterns of behavior in which each

Preprint submitted to 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. Received March 22, 2013.

interactant's actions influence the other's behavior [9]. In this area a number of features of inter-human social behavior stand out. Timing, turn-taking, and synchronization dynamics in dyadic interaction has long been recognized as fundamental for communication [10]. Developmental psychologists have proposed that an important aspect of satisfactory positive interaction (for instance during social play) is reciprocal involvement, expressed by the level of mutual responsiveness as observed in conversational turn-taking [11]. Recently, several authors have highlighted the impact of these dynamics of interaction, like the timing and facial/gestural expressiveness, in the study of human-robot interaction [12-14].

To investigate such effects, Staudte & Crocker [12] exposed subjects to videos of a robot gazing at different objects in a linear array tangent to the line of sight, and sentences referring to these objects, that were either congruent or incongruent with the videos. Subjects were to respond whether the sentence accurately described the scene. The principal findings of the study is the effect of robot gaze on human performance, with most rapid performance for congruent gaze, poorest performance for incongruent gaze, and intermediate performance when the robot made no gaze. Extending such studies into the domain of actual physical HRI, Huang et al. demonstrated that when task-related gaze cues anticipate the linguistic references in verbal communication, recall and response times are significantly improved vs. conditions where gaze is delayed or inconsistent [13]. Similarly, Boucher et al. demonstrated that when robot gaze is directed to a target object prior to the completion of the verbal specification of that object, subjects can anticipate the spoken specification, and begin to manipulate that object with significantly reduced (even negative/anticipatory) reaction times, vs. conditions where gaze is masked or eliminated [14]. These studies indicate the crucial role of gaze in coordinating joint action.

In the behavioral sciences context of joint action, it is often difficult to determine whether an activity that is performed by multiple agents should be conceived of as a joint collaborative activity that is based upon joint intentions or merely as the common outcome of individual intentions. For example, each individual agent might be acting on an individual intention towards an individual goal, and even though the outcome emerges from the combined efforts of the agents, they are not necessarily acting jointly. In other words, what qualifies as a plural activity [15] does not necessarily qualify as a joint collaborative activity. In order to test whether coordinated joint activity can emerge in the absence of joint plans, one could experimentally manipulate the presence/absence of joint plans in experimenters who would interact with naïve subjects. However, while manipulating a human experimenter to use, or not, a joint plan is methodologically difficult if not impossible, this it is technically feasible with a robot subject.

The motivation for the current research is thus twofold. The principal motivation is to manipulate the presence vs. absence of a joint plan within the robot cognitive system, in order to determine if joint task outcome with naïve subjects will be improved in the presence of a joint plan vs. parallel individual plans. The second motivation, within this context, is to determine the effects of spoken language and task relevant gaze cues in the successful communication and achievement of joint action.

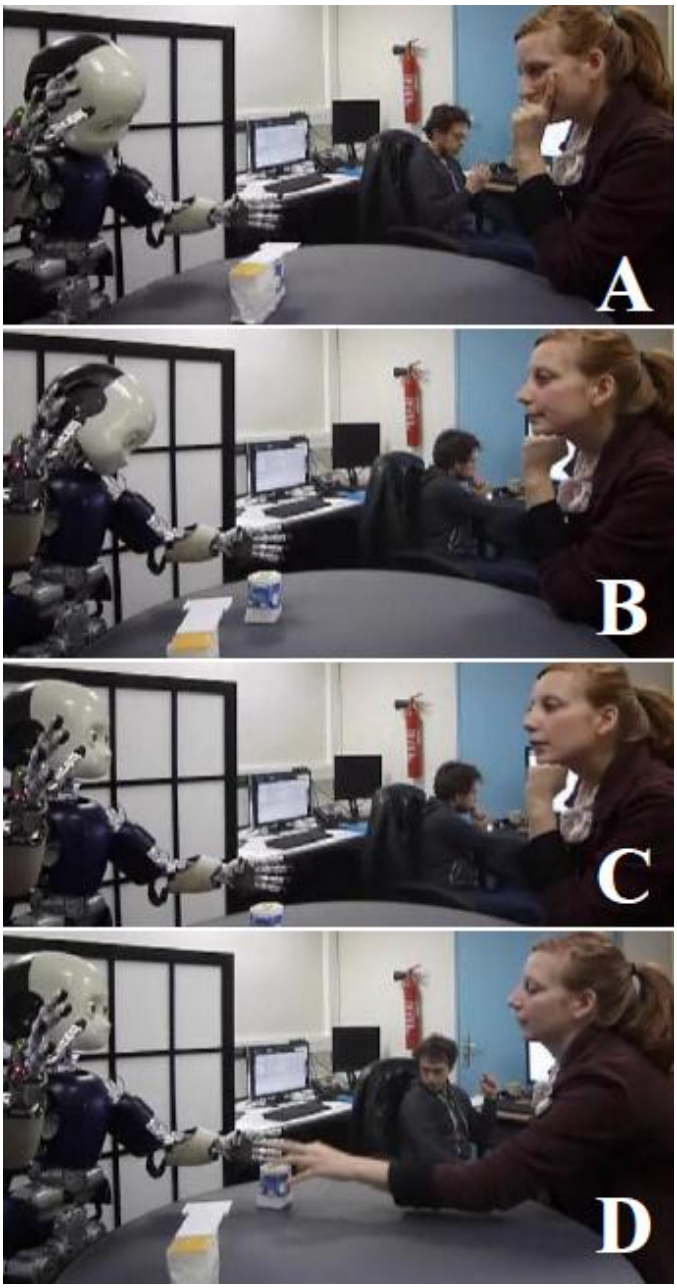

Fig. 1. Cooperative interaction task, in conditions with Joint plan, communicated by gaze alone, without speech. A. Yellow box initially covers blue toy. Human observes as robot moves towards box to uncover toy. B. Robot gazes to target object for human to grasp. C. Robot looks to Human to indicate that human should act. D. Human responds to gaze cue and initiates action to move the toy to the indicated location.

We will report on experiments in which twelve naïve subjects interact in a cooperative task with the iCub humanoid robot [16] under four experimental conditions that involve full cooperation, cooperation with no speech, cooperation with no

gaze, and finally the "solo" condition in which the robot does not use a joint plan. We measure the effects of these manipulations on several specific measures of cooperation performance. The human and robot are to work together to achieve the goal of uncovering a toy with a box, so that the exposed toy can then be retrieved, illustrated in Figure 1.

II. COOPERATIVE ROBOT SYSTEM METHODS

The experiments were performed with the iCub robot [16], and the ReacTable™ instrumented table that could detect the identity and location of objects placed upon it, illustrated in Figure 1. The iCub is controlled by a cognitive system [6-8] that provides for the creation and use of joint plans for the execution of sensory-guided actions in cooperation with the human subject. Our novel contribution with respect to cognitive system development is the introduction of coordinated gaze, speech and shared plan learning and execution, illustrated in Figure 3 and developed in section B below.

A. Cognitive System for Cooperation

The cognitive system is based on the creation and use of joint plans. A joint plan is defined as a sequence of actions with each action allocated to one of the agents, such that both agents represent this plan, and use it to achieve their shared goal. A hallmark of the joint plan is that it allows for role reversal – that is – the cooperation partners can reverse their complementary roles, with each taking on the previous role of the other, respectively [17].

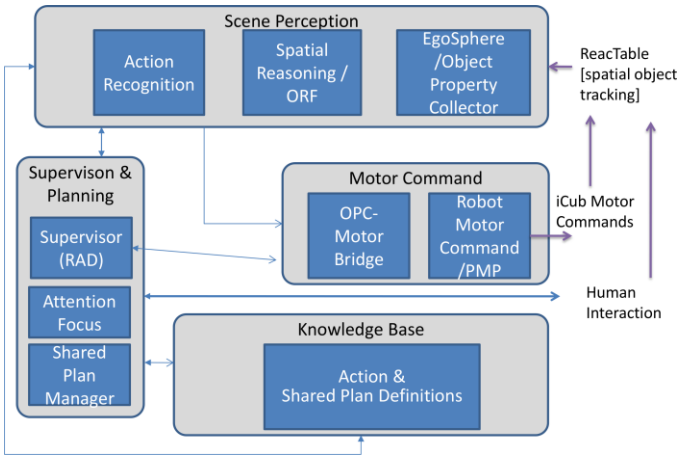

Fig. 2. Control Architecture for cooperative interaction. Spatial location of objects on ReacTable communicated to *Object Property Collector* which stores the state of the world. *Spatial reasoning* detects object movement and spatial relations among them providing the required information for *Action Recognition*. *Supervision and Planning* monitors interaction, focuses gaze attention on linguistic references, and manages joint/shared plan execution. Shared plans and action definitions are stored for long term use in *Knowledge Base*. *Motor command* monitors manipulation of object predicates transformed to motor space using passive motor paradigm (PMP). See text for details.

The current research thus exploits a series of developments that have resulted in the ability of the robot to learn a joint plan, to use that joint plan in cooperation with a human, and to demonstrate role reversal [6-8]. Again, in role reversal, the cognitive agent is able to use the same joint plan, but exchange roles with the other partner. This ability has been recognized by developmental psychologists as evidence that the agent has "bird's eye view" knowledge of the joint plan, rather than a purely ego-centric view [17].

The system is outlined in Figure 2. Joint/shared plans are managed by the Supervision and Planning subsystem. Through spoken language interaction, new joint plans can be established through different combinations of spoken language specification, imitation or demonstration. Spoken language generated by the robot has a semantics that is defined in terms of the shared task [18]. Verbs refer to the actions of manipulating objects on the table, common nouns refer to objects and pronouns to the robot and human. In the current experiment, a pre-learned joint plan for the shared task is employed. The finality of this plan is that the initially covered toy is uncovered and retrieved. The first agent uncovers the toy. The second agent can then retrieve the toy. Finally the first agent replaces the box on the table.

Perceptual information about the location of objects and the human partner is extracted from the ReacTable™ (see below), and stored in the Object Properties Collector (OPC). When the joint plan calls for the robot to manipulate an object, the spatial coordinates of that object in the task-space of the robot are used to generate the appropriate action (specified in more detail below).

In order to coordinate the unfolding joint action, the robot must perceive when the human has performed his actions. A simple spatial reasoning engine detects spatial relations of proximity and change in position based on OPC, so that Action Recognition can detect actions including *put(object, location)* [6].

B. Coordinated Speech and Gaze Control

In order to coordinate speech and gaze, the Attention Focus module of Supervision and Planning handles the translation of actions in the joint plan, and determines the linguistic references. When the linguistic referent in the utterance is identified the gaze is directed there, and the word is sent to the speech synthesizer. This results in continuous speech with coordinated gaze. Gaze thus attains the target several hundred milliseconds before the linguistic reference is pronounced. At each step in the unfolding of a plan (joint or solo), the robot retrieves the current action, and identifies the agent, the object and the final location where that object should be placed. In conditions where gaze is active, prior to the execution of the next action, the robot directs coordinated gaze movements to the agent (if the agent is the human), then the objet, and finally the desired location where that object should be placed. In conditions where language communication is active, the gaze is coordinated with the timing of the spoken language. That is, if the robot says "You put the box on the left", as each word you, box and left is spoken, the gaze is directed to that location. The location of the human subject is pre-specified based on the experimental set up as illustrated in Figure 1. The temporal structure of speech-gaze coordination is illustrated in Figure 3.

Preprint submitted to 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. Received March 22, 2013.

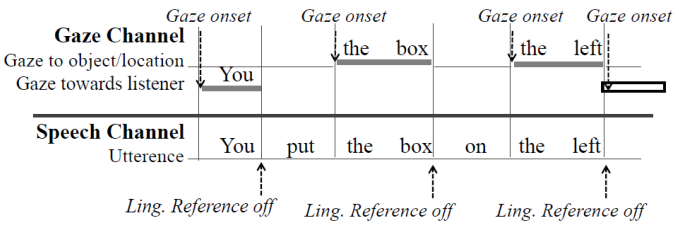

Fig. 3. Speech and gaze coordination. For all pertinent linguistic references (agent, object, recipient location), gaze is directed to the referent object, location, or human just prior to pronunciation of the linguistic referent. The period where gaze precedes linguistic referent termination is indicated by thick grey lines in the Gaze Channel. Final imperative gaze to user indicated in unfilled line on "Gaze towards listener." When present, gaze thus provides an additional communication channel for joint plan management. Figure format modified from [13].

To achieve this coordination, the iCub eyes can move independently in the horizontal and vertical head centered orbits, and the head has three additional degrees of freedom. Coordinated gaze, as eye-head motion can be directed with an inverse kinematics engine that will take the eyes to a target in the three-dimensional space task space surrounding in the iCub. These movements are coordinated with an initial oculomotor saccade that is then followed by a slower head movement. The robot's eye movement and head movement completion times are respectively 100ms and 600ms. The eye thus attains the target first, with the head continuing to move to the target, and the eyes compensating for this continued head movement in order to stay fixated on the target. The generation of these human-like movements studied in human gaze is achieved in the robot with the iCub gaze controller. The controller employed to coordinate the iCub gaze acts intrinsically in the Cartesian space, taking as input the spatial location of the object of interest where to direct the robot attention, and then generates proper minimum-jerk velocity commands simultaneously to the neck and the eyes.

C. Spatial Localization and Accurate Object Manipulation

Objects are manipulated by the human and robot on an instrumented table (ReacTable™), as illustrated in Figure 1. Fiducial markers on the base of objects are detected by an IR camera beneath the ReacTable surface, providing millimeter accuracy for object localization. The 2D surface of the table is calibrated into the joint space of the iCub by a linear transformation calculated based on a sampling of three calibration points on the table surface that are pointed to by the iCub. Thus, three points are physically identified in the Cartesian space of the iCub, and on the surface of the ReacTable, thus providing the basis for calculation of a transformation matrix which allows the projection of object coordinates in the space of the table into the Cartesian space of the iCub. These coordinates can then be used as spatial arguments to the motor control system of the iCub.

Motor control is provided by PMP. The Passive Motion Paradigm (PMP) [19] is based on the idea of employing virtual force fields in order to perform reaching tasks while avoiding obstacles, taking inspiration from theories conceived by Khatib during 80's [20]. with a tool that relies on a powerful and fast nonlinear optimizer, namely Ipopt [21]; the latter manages to solve the inverse problem while dealing with constraints that can be effectively expressed both in the robot's configuration space (e.g. joints limits) and in its task-space. This new tool [22] represents the backbone of the Cartesian Interface, the software component that allows controlling the iCub directly in the operational space, preventing the robot from getting stuck in kinematic singularities and providing trajectories that are much smoother than the profiles yielded by the first implementation of PMP.

III. HUMAN EXPERIMENTAL METHODS

A total of N = 12 naïve university subjects were tested in each of four conditions (specified in Table 1) in a human-robot cooperation task, illustrated in Figure 1. The task was based on experimental paradigms used with human infants [4]. Specifically, the goal of the shared task was to retrieve a small object, the "toy" that was covered by a larger object, the "box". One participant would lift the box, allowing the other to take the toy, and finally the first participant would replace the box on the table. Thus the joint plan requires three successive movements, allocated as stated to the two participants.

Prior to the start of the experiments, subjects were informed of the structure of the task, and then were shown an example of how the shared task unfolded, with one of the experimenters interacting with the robot. Subjects were simply instructed to attempt to achieve the joint goal of retrieving the hidden toy with the robot.

Four conditions were tested, which manipulated the use of a joint plan. Each subject was exposed to *4 repetitions of each of the 4 conditions*, twice initiating and twice moving second, for a total of 16 trials per subject. The order of conditions and who starts the interaction were pseudo-randomized across subjects in order to balance across all conditions.

	Functions		
Conditions	Joint plan	Spoken Communication	Gaze
Full	x	x	x
No Language	x		x
No Gaze	x	x	
Solo			x

Table 1. Specification of experimental conditions as a function of the cognitive/communicative capabilities that were activated. this is a 3 (gaze, language, both) x 2 (shared plan, solo) design with two conditions removed from the solo case. The two missing solo-language conditions were removed because language would have made it explicit that it was a solo run. This would have explicitly prevented subjects from turn taking, and would thus have biased the experiment.

The *Full* condition corresponded to the full cooperation capability that we had developed for optimal cooperative human-robot activity. This included the use of a joint plan, which specifies the successive, interlaced, actions of the robot and human; spoken language communication, whereby the robot announces at each step who does what; and gaze, whereby the gaze of the robot is directed to the human (when it is his turn) then the target object to manipulate, and then the destination location where that object should be placed.

The *No Language* condition was identical, with the exception that there was no spoken communication. The *No Gaze* condition was identical to the Full condition, with the exception that the gaze remained fixed throughout each trial. Finally, in the solo condition, there was no joint plan, no spoken communication, and gaze is only directed to the object and target locations for each movement.

IV. Results

Results were quantified in terms of the following dependent variables. (1) Cooperation quality, as indicated by turn-taking quality (see below), (2) Number of collisions, (3) Number of attempts the participant wanted to assist the robot (either "cleaning up" the robot's mistake or trying to prevent mistakes by the robot), (4) The number of cognitive errors performed by the participants. These variables were assessed by a trained behavioral scoring expert at the Max Plank Institute (KH) by analysis of high resolution films of the interactions that were recorded at the Robot Cognition Laboratory. Standard blind methods for coding of behavior were used. After verifying the normal distribution of the data, all statistical analyses are performed using one way ANOVAs and Scheffe post-hoc comparisons. One trial corresponds to one complete execution of the joint task of uncovering and retrieving the toy.

A. Cooperation

Results are presented in terms of successful cooperation, on a scale from 0 – 3. Cooperation 3 corresponds to perfect turn taking, with sequences of actions HRH, RHR. Cooperation 2 corresponds to some turn taking, even though not perfectly alternating, e.g. HHR, RRH. Cooperation 1 corresponds to no turn taking, but all actions are carried out by one of the agents, e.g. HHH, RRR. Finally, Cooperation 0 corresponds to no turn taking, and not all actions being carried out, e.g. HH, H, R, RR.

The results in Figure 4 clearly indicate that cooperation is impaired in the solo condition, i.e. that the presence of a joint plan in the robot yields significantly better cooperation. This was confirmed by the significant Condition effect, $F(3, 33) = 15.16$, $p < 0.0001$. Post hoc comparisons confirmed that Solo cooperation was significantly reduced compared to the other three conditions. No other comparison was significant. In order to determine if at least some successful instances of cooperation (i.e. fully completing the shared task, with three successive alternating actions) could occur in the absence of a joint plan, we also examined the percentage of different levels of cooperation performance by condition, as illustrated in Figure 5. Here we can see that although the percentage of level 3 cooperation was reduced in the solo condition, more than 30% of the interactions were successfully completed in the solo condition (i.e. instances of Cooperation 3 level performance). This indicates that behavior that can appear to be cooperative to an external observer can be achieved, even though at least one of the partners (here the robot) is not using a joint plan.

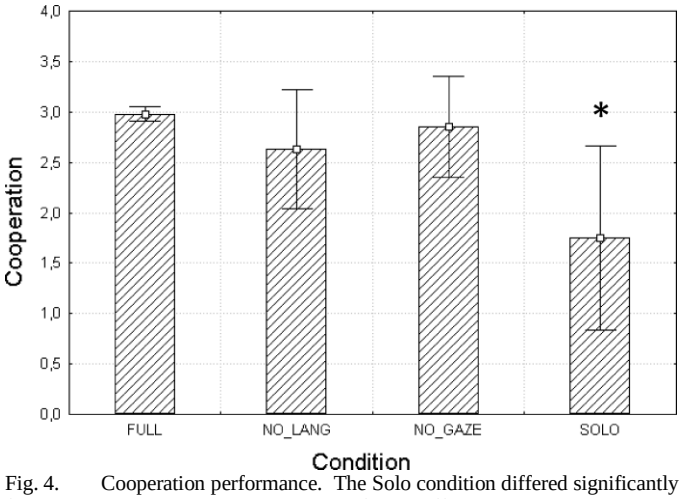

Fig. 4. Cooperation performance. The Solo condition differed significantly from all other conditions. No other significant differences were obtained: the three joint plan conditions were similar in their establishment of turn-taking.

However, all other conditions yielded greater performance. We also observed that all 12 subjects were able to achieve successful completion of the task in conditions that include the joint plan, whereas only 5 did so in the solo condition.

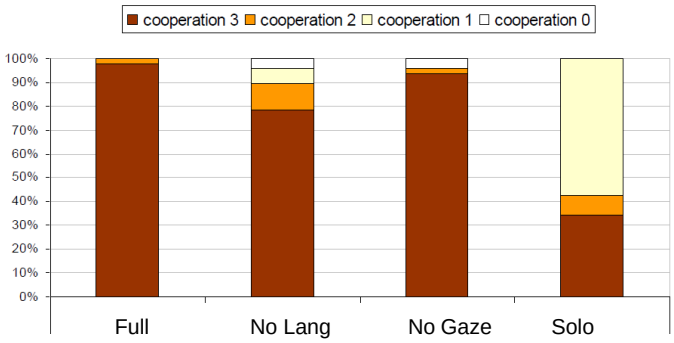

Fig. 5. Distribution of different levels of cooperation across conditions, in percentage of trials.

B. Collisions

Collisions correspond to interruption of actions, when the human subject starts to perform an action but withdraws his hand due to the robot's intervening movements. Figure 6 indicates that the solo and no language conditions lead to similarly high numbers of collisions. This was confirmed as the effect of condition was significant, $F(3,33)=4.93$, $p< 0.01$. Planned comparisons revealed that Full and No Gaze conditions resulted in significantly less collisions that the No Language and Solo conditions, thus highlighting the importance of language in avoiding collisions.

.

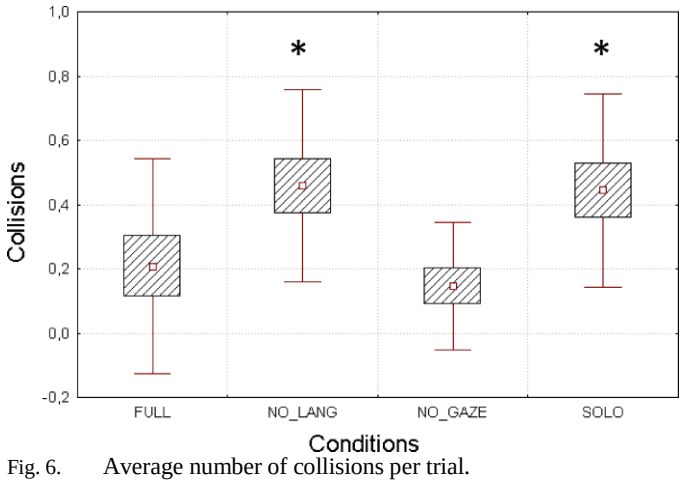

Fig. 6. Average number of collisions per trial.

C. Human Helping the Robot

Each trial of the cooperative interaction involved several actions to be performed by the robot. If the robot goes first, it will pick up the box and move it, then let the human take the toy and place it on the table, and finally replace the box in the central location. If the robot goes second, then these roles are reversed. For each movement of the robot, the human can assist the robot by either helping it in making the initial grasp, or in positioning the object correctly after the robot has placed it.

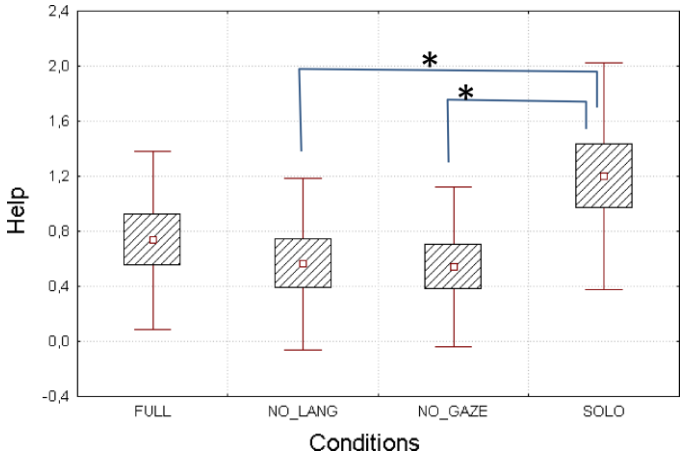

Fig. 7. Average number of instances of assistance per trial.

In Figure 7 we see that there was significantly more assistance by the subject in the solo condition than in other conditions. This is substantiated by the ANOVA, $F(3,33)=6.80$, $p<0.005$. Post-hoc (Scheffe) tests revealed that subjects help the robot significantly more in the Solo vs. No Lang and No Gaze conditions ($p < 0.01$). No other effects were significant. That the naïve subjects frequently helped the robot when it was in solo mode indicates that they were involved in the task, and indeed demonstrated a form of mutual responsiveness and commitment to the shared goal. In addition, it indicates that the success in the solo condition was partly due to the subjects' intervention, rather than an error-free interaction between subject and robot, providing further evidence that the joint plan and subjects' commitment to the shared goal lead to superior performance.

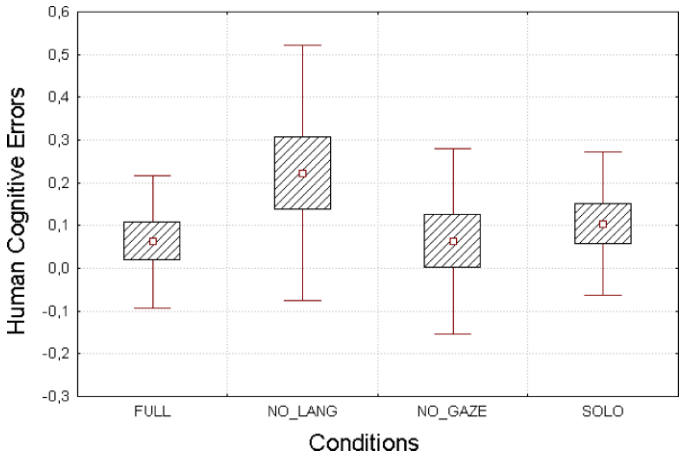

Fig. 8. Average number of cognitive errors per trial.

D. Human Cognitive Errors

Human participants sometimes performed wrong actions, i.e. used the wrong object, put the correct object at the wrong place, or took turns when it was the robot's turn. The scores in Figure 8 represent average numbers of cognitive errors per trial. We can observe a small increase in errors in the no language condition, though the ANOVA reveals no effect, $F(3,33) = 1.6$, $p = 0.19$.

Participants tended to make more errors in the no language condition as compared to the full condition. Post hoc comparison revealed that there was a trend towards this effect ($p = .061$). No other effects approached significance. These results suggest that in the absence of spoken communication, the naïve subjects could not always anticipate all necessary steps in the sequence from the objects and nonverbal cues alone, though the statistics are not conclusive.

V. Discussion

This research can be situated in the context of joint planning and collaboration [23]. Part of its novelty results from the successful collaboration between human developmental science [2] and cognitive robotics [6-8]. The experimental protocol was developed to test the hypothesis that while behavior resembling cooperation (including coordinated alternating action towards a final goal) can be achieved without an explicit joint plan, the actual use of a joint plan will result in better cooperative behavior. Testing such a hypothesis with human experimenters is difficult or impossible, as it requires that the experimenter carefully control their gaze, speech, and timing of actions in a controlled and repetitive manner. In contrast, such manipulations are ideal for robot interaction scenarios, as the behavior of the robot can be controlled in a standard and repetitive manner.

Preprint submitted to 2013 IEEE/RSJ International Conference
on Intelligent Robots and Systems. Received March 22, 2013.

We confirmed that human subjects performed best when the robot was fully cooperative, using the joint plan, and communicated the joint plan both using gaze and speech. Subjects also performed well in the condition (no gaze) where the joint plan was used, and communicated by speech alone. This indicates that speech is a potent modality for the on-going maintenance of cooperative interaction.

It was striking that robot and human cooperated successfully even in the absence of verbal communication. Specifically, a crucial aspect of successful performance is to correctly determine who goes first. In the conditions with spoken communication, this is relatively easy as the robot announces who should do what at each step. However, subjects and robots were able to achieve successful collaboration based upon nonverbal communication alone, with subjects closely following the robot's gaze and being able to interpret who should go first. Importantly, this only occurred in conditions in which the robot represented the task as involving a joint plan in which another agent performs complementary roles. Thus, nonverbal communication often seems to be sufficient to coordinate action roles between robots and naïve subjects, but it critically depends on the robots representation of a joint plan rather than the physical sequence of events alone.

In the no language condition, the robot directed its gaze at the human when it was the human's turn to move – including when the human should start the trial. Data presented in Figures 3 and 4 indicates that the in the absence of speech communication, humans followed the robot's gaze, and correctly interpreted the human-directed gaze as an invitation to start the trial.

Gaze indeed plays a central role in the real-time orchestration of human interaction. It has been demonstrated that in conditions where verbal instructions are ambiguous, the speaker disambiguates first by gaze and then via language, and the listener can use this unambiguous gaze prior to the availability of the disambiguating language [24]. Such gaze cues should and can be exploited in the domain of human-robot interaction. Huang et al. [13] demonstrated that the use of human-like gaze in human-robot interaction resulted in improved memory-recall, quality of collaborative work, and even the human perception of the robot. In the collaborative task, the robot indicated where to place different lego blocks in a categorization task. As in our manipulation, when the target object was mentioned, the gaze was directed to that object before the end of the utterance. Indeed, these authors initiated the gaze prior to the speech onset.

In the current research we have exploited this use of task-related gaze in order to address a question concerning the status of the joint plan in coordinated activity that is performed by multiple agents [15]. Our results argue for the "cooperative stance" which holds that joint action is most successful in the presence of a true joint plan, consistent with a current line of developmental research [1-5].

Interestingly, these data also indicate that while the solo condition yielded perturbed performance, there were also cases with successful turn-taking behavior, providing support for the argument that places less emphasis on the necessity of a full blown joint plan [15]. When the human "jumped in" to the plan, it was only in the solo condition that the robot did react adequately by just doing the next necessary thing; in all the other conditions, this capacity was not available, as the robot expected the human to follow the joint plan. Conceptually, this argues that perhaps the best condition will be a more flexible joint plan, capable of adapting to deviations from the canonical plan that could take place during execution. In other words, one might speculate that a system that is capable of performing both an individual and a joint plan of action might be most versatile: the default state is to expect the cooperative partner to take her turn, but switch over to an individual plan when the partner fails to make a move.

In conclusion, we demonstrate that a robot that has a complete plan for a sequence of actions directed towards a target final state can produce turn taking behavior that resembles true cooperative activity. However, we also observe that this cooperation capability is significantly enhanced when the robot has a joint plan which allows it to guide the successive turn taking in achieving the execution that ends in the final target state.

This research demonstrates that humanoid robots can be used with naïve subjects in the testing of human behavior that requires the precise manipulation of behavioral parameters including the use of shared vs. solo plans, speech and task-oriented gaze. The reliable manipulation of these inherent social interaction parameters in humans is difficult or impossible, thus the use of social robots provides a new testing ground for such research. In doing so, this research also contributes in a concrete and specific manner to the identification of behavioral capabilities that will contribute to more robust cooperative human-robot interaction. In particular, the combined use of joint plans that can be modified at execution time, communicative speech, and task-oriented gaze have been demonstrated here to contribute to robust, adaptive, joint activity in human robot cooperation.

ACKNOWLEDGMENT

This research has been funded by the European Commission under grants EFAA (ICT-270490) and CHRIS (ICT-215805).

REFERENCES

[1] K. Hamann, F. Warneken, and M. Tomasello, "Children's developing commitments to joint goals," *Child Dev*, vol. 83, pp. 137-45, Jan-Feb 2012.

[2] K. Hamann, F. Warneken, J. R. Greenberg, and M. Tomasello, "Collaboration encourages equal sharing in children but not in chimpanzees," *Nature*, vol. 476, pp. 328-31, Aug 18 2011.

[3] M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll, "Understanding and sharing intentions: The origins of cultural cognition," *Behavioral and Brain Sciences*, vol. 28, pp. 675-691, 2005.

[4] F. Warneken, F. Chen, and M. Tomasello, "Cooperative activities in young children and chimpanzees," *Child Development*, vol. 77, pp. 640-663, 2006.

[5] F. Warneken and M. Tomasello, "Helping and cooperation at 14 months of age," *Infancy*, vol. 11, pp. 271-294, 2007.

[6] S. Lallée, S. Lemaignan, A. Lenz, C. Melhuish, L. Natale, S. Skachek, T. van Der Tanz, F. Warneken, and P. Dominey, "Towards a Platform-Independent Cooperative Human-Robot Interaction System: I. Perception," in *IROS*, Taipei, 2010.

[7] S. Lallée, U. Pattacini, J. Boucher, S. Lemaignan, A. Lenz, C. Melhuish, L. Natale, S. Skachek, K. Hamann, J. Steinwender, E. A. Sisbot, G. Metta, R. Alami, M. Warnier, J. Guitton, F. Warneken, and P. F. Dominey, "Towards a Platform-Independent Cooperative Human-Robot Interaction System: II. Perception, Execution and Imitation of Goal Directed Actions," in *IROS*, San Francisco, 2011, pp. 2895 - 2902.

[8] S. Lallée, U. Pattacini, S. Lemaignan, A. Lenz, C. Melhuish, L. Natale, S. Skachek, K. Hamann, J. Steinwender, E. A. Sisbot, G. Metta, J. Guitton, R. Alami, M. Warnier, T. Pipe, F. Warneken, and P. Dominey, "Towards a Platform-Independent Cooperative Human-Robot Interaction System: III. An Architecture for Learning and Executing Actions and Shared Plans," *IEEE Transactions on Autonomous Mental Development*, vol. In press, 2012.

[9] J. K. Burgoon, L. A. Stern, and L. Dillman, *Interpersonal adaptation: Dyadic interaction patterns*: Cambridge University Press, 2007.

[10] J. Nadel, *New perspectives in early communicative development*: Routledge, 1993.

[11] T. Stivers, N. J. Enfield, P. Brown, C. Englert, M. Hayashi, T. Heinemann, G. Hoymann, F. Rossano, J. P. de Ruiter, K. E. Yoon, and S. C. Levinson, "Universals and cultural variation in turn-taking in conversation," *Proc Natl Acad Sci U S A*, vol. 106, pp. 10587-92, Jun 30 2009.

[12] M. Staudte and M. W. Crocker, "Visual attention in spoken human-robot interaction," presented at the Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, La Jolla, California, USA, 2009.

[13] C.-M. Huang and B. Mutlu, "Robot Behavior Toolkit: Generating Effective Social Behaviors for Robots. ," in *7th ACM/IEEE Conference on Human-Robot Interaction (HRI 2012)*, Boston, MA., 2012.

[14] J. D. Boucher, U. Pattacini, A. Lelong, G. Bailly, F. Elisei, S. Fagel, P. F. Dominey, and J. Ventre-Dominey, "I Reach Faster When I See You Look: Gaze Effects in Human-Human and Human-Robot Face-to-Face Cooperation," *Front Neurorobot*, vol. 6, p. 3, 2012.

[15] S. Butterfill and N. Sebanz, "Editorial: Joint Action: What Is Shared?," *Review of Philosophy and Psychology*, vol. 2, pp. 137-146, 2011.

[16] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. von Hofsten, K. Rosander, J. Santos-Victor, A. Bernardino, and L. Montesano, "The iCub Humanoid Robot: An Open-Systems Platform for Research in Cognitive Development," *Neural Networks, Special issue on Social Cognition: From Babies to Robots*, vol. 23, 2010.

[17] M. Carpenter, M. Tomasello, and T. Striano, "Role reversal imitation and language in typically developing infants and children with autism," *Infancy*, vol. 8, pp. 253-278, 2005.

[18] I. A. Smith and P. R. Cohen, "Toward a semantics for an agent communications language speech-acts," in *Thirteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference*, 1996.

[19] V. Mohan, P. Morasso, G. Metta, and G. Sandini, "A biomimetic, force-field based computational model for motion planning and bimanual coordination in humanoid robots," *Autonomous Robots*, vol. 27, pp. 291-307, 2009.

[20] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," *Int. J. Rob. Res.*, pp. 90-98, 1986.

[21] A. Wätcher and L. T. Biegler, "On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming," *Mathematical Programming*, vol. 106, pp. 25-57, 2006.

[22] U. Pattacini, F. Nori, L. Natale, G. Metta, and G. Sandini, "An Experimental Evaluation of a Novel Minimum-Jerk Cartesian Controller for Humanoid Robots," in *IROS*, Taipei, 2010.

[23] C. Breazeal, G. Hoffman, and A. Lockerd, "Teaching and working with robots as a collaboration," in *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 3*, 2004, pp. 1030-1037.

[24] J. Hanna and S. Brennan, "Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation," *Memory & Language*, vol. 57, p. 21, 2007.

Preprint submitted to 2013 IEEE/RSJ International Conference
on Intelligent Robots and Systems. Received March 22, 2013.