

# Effects of anticipatory perceptual simulation on practiced human-robot tasks

Guy Hoffman · Cynthia Breazeal

Received: 6 February 2009 / Accepted: 7 December 2009 / Published online: 19 December 2009  
© Springer Science+Business Media, LLC 2009

**Abstract** With the aim of attaining increased fluency and efficiency in human-robot teams, we have developed a cognitive architecture for robotic teammates based on the neuropsychological principles of anticipation and perceptual simulation through top-down biasing. An instantiation of this architecture was implemented on a non-anthropomorphic robotic lamp, performing a repetitive human-robot collaborative task.

In a human-subject study in which the robot works on a joint task with untrained subjects, we find our approach to be significantly more efficient and fluent than in a comparable system without anticipatory perceptual simulation. We also show the robot and the human to improve their relative contribution at a similar rate, possibly playing a part in the human's "like-me" perception of the robot.

In self-report, we find significant differences between the two conditions in the sense of team fluency, the team's improvement over time, the robot's contribution to the efficiency and fluency, the robot's intelligence, and in the robot's adaptation to the task. We also find differences in verbal attitudes towards the robot: most notably, subjects working with the anticipatory robot attribute more human qualities to the robot, such as gender and intelligence, as well as credit for success, but we also find increased self-blame and self-deprecation in these subjects' responses.

---

**Electronic supplementary material** The online version of this article (<http://dx.doi.org/10.1007/s10514-009-9166-3>) contains supplementary material, which is available to authorized users.

---

G. Hoffman (✉) · C. Breazeal  
MIT Media Laboratory, 20 Ames Street E15-468, Cambridge,  
MA 02142, USA  
e-mail: [guy@media.mit.edu](mailto:guy@media.mit.edu)

C. Breazeal  
e-mail: [cynthiab@media.mit.edu](mailto:cynthiab@media.mit.edu)

We believe that this work lays the foundation towards modeling and evaluating artificial practice for robots working in collaboration with humans.

**Keywords** Human-robot interaction · Perceptual simulation · Anticipation · Human-robot teamwork · Joint practice · Human-subject studies · Cognitive models · Top-down bias · Priming

## 1 Introduction

Our goal is to design robots that can work fluently with a human partner in a physically situated setting. *Fluency* in joint action is the quality existent when two agents perform together at high level of coordination and adaptation, in particular when they practice a task repetitively, and are well-accustomed to the task and to each other. This quality is observed in a variety of human behaviors, but is virtually absent in human-robot interaction.

Neurological and psychological evidence in humans indicates that anticipation and perceptual simulation (the internally-originating activation of perceptual neural pathways) plays a role in perception, in the perception of conspecifics, and in joint action (Wilson and Knoblich 2005; Sebanz et al. 2006). In simulated agents acting with humans, we have shown anticipation to lead to improved task efficiency and fluency, as well as a perceived commitment of a simulated robot to the team and its contribution to the team's fluency and success (Hoffman and Breazeal 2007).

Based on these findings, we believe that anticipation through perceptual simulation can provide a powerful model for robots acting jointly with humans if they are to collaborate fluently using multi-modal sensor data. To that end,

we developed a cognitive architecture based on the principles of embodied cognition and top-down perceptual simulation, ideas which are gaining ground in the neuroscientific literature in recent years (Barsalou 1999; Spivey et al. 2005; Wilson 2002); for a review, see: Hoffman and Breazeal (2006).

In this paper we introduce some core concepts of our cognitive framework and its implementation on a non-anthropomorphic robot designed for human-robot collaboration. We discuss a controlled human subject study conducted to evaluate the performance of the implemented system, and the effects it has on the efficiency and fluency of the task, as well as on the human subjects' perception of the robot and the team. We are particularly interested in how the system performs within the context of practice, in which the human and the robot repeat a set of identical actions.

### 1.1 Related work

Human-robot collaboration has been investigated in a number of previous works, although the question of fluent action meshing or the improvement thereof through repetition or practice has not received much attention. Kimura et al. (1999) have studied a robotic arm assisting a human in an assembly task. Their work addressed issues of vision and task representation, but does not address anticipation, fluency, or practice. Other human-robot collaboration work, such as that of Fong et al. (2001) or Jones and Rock (2002) studies human-robot collaboration with an emphasis on dialog and control, aimed primarily at the teleoperation scenario.

Some work in shared-location human-robot collaboration has been concerned with the mechanical coordination and safety considerations of robots in shared tasks with humans (Woern and Laengle 2000; Khatib et al. 2004). Other work addresses turn-taking and joint plans, but not anticipatory action, practice, or fluency (Hoffman and Breazeal 2004). Anticipatory action, without relation to a human collaborator has been investigated in the area of robot navigation, e.g. Endo (2005).

The idea of top-down biasing has been utilized in computational systems in the past, for example in visual action recognition (Bregler 1997). Wren and Pentland created a robust human dynamic recognition and classification system by feeding likelihood data from high-level HMM procedures to pixel-level classifiers (Wren et al. 2000). Similarly, Hamdan et al. (1999) classified gesture sequences using Continuous Density Hidden Markov Models. Ude et al. (2007) discuss similar top-down processing ideas for visual attention on a humanoid robot. None of these works, however, model the top-down influences as perceptual simulation using the same pathways used for bottom-up processing, as is supported by the neuro-psychological literature, and proposed in this paper.

Some neurologically-inspired agent systems address the dichotomy between fast and slow action-generation. For example, Marsella and Gratch (2009) model emotional appraisal as divided into rapid "perceptual" processing and slower "inferential" processes. Similarly, Duffy (2000) proposes a Social Robot Architecture, which distinguishes between "reactive" and "deliberative" mechanisms. Both these and similar approaches focus on the dynamics between two tiers of reactive behaviors, but describe the two kinds of systems as distinct and separate, albeit interrelated. Such symbolic approaches do not model anticipation in a way that takes into consideration the gradual transition from deliberate to automatic behavior as it occurs in repetitive perception, as we propose in this work. Moreover, our system focuses on the perceptual pathways, introducing the notion of combining real perception with simulated perception as a mechanism of improved reaction times for collaborative robots.

Our own previous work in anticipatory action to support human-robot fluency was implemented on a simulated agent, using a discretized model framed as a stepwise MDP with simulated perception, and no perceptual simulation (Hoffman and Breazeal 2007). This paper significantly extends this work as it models anticipation through the simulation of perceptual symbols. Furthermore, the work presented here is implemented on a physical robot using noisy, continuous sensory input, acting in a situated interaction with a moving human.

## 2 Cognitive architecture

This section provides a brief outline of the principles guiding the cognitive architecture employed in the studies described below. It is by no means complete, and primarily serves to set the stage for the experimental results obtained in the human subject study. A full description of the cognitive architecture is the topic of a separate publication.

We propose that fluency in joint action achieved through practice rests on two premises: (a) *anticipation* based on a model of repetitive past events, and (b) the modeling of the resulting anticipatory expectation as *perceptual simulation*, affecting a top-down bias of perceptual processes.

To allow for this approach, we model concepts leading to actions not as amodal symbolic structures, but as instances of activations residing within the perceptual streams that process sensory data. Thus, actions are triggered bottom-up through activation originating in perceptual stimulation, and conversely anticipated concepts bias the sensory pathway detecting the features of that concept, leading to diminished reaction times for confirmatory sensory events, and resulting in higher fluency and efficiency in practiced actions.

## 2.1 Modality streams and process nodes

In our model, sensory input is processed in *modality streams* built of interconnected *processing nodes* (Fig. 1). These nodes can correspond to raw sensory input (such as a visual frame or a joint sensor), to a feature (such as the dominant color or orientation of a sensory data point), to a property (such as the speed of an object), or to a higher-level concept describing a statistical congruency of features of properties. This structure is in the spirit of the Convergence Zones described in Simmons and Barsalou (2003).

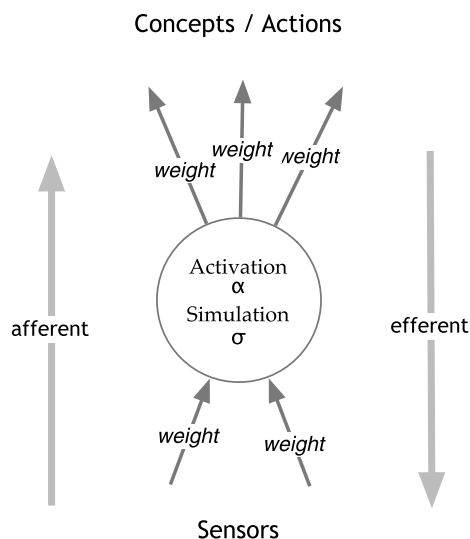
While there is no inherent difference between these categories, in our work we often think about different layers of nodes, which we refer to—from bottom to top—as: *sensor*, *feature*, *property*, and *concept* nodes.

Modality streams are connected to an *action network* consisting of *action nodes*, which are activated in a similar manner as perceptual processing nodes. An action node, in turn, leads to the performance of a motor action, and is usually fed by the Concept node layer.

Connections between nodes in a stream are not binary, but weighted according to the relative influence they exert on each other.

Importantly, activation flows in both directions, the *afferent*—from the sensory system to concepts and actions—and the *efferent*, in the opposite direction. This affordance of top-down (efferent) perceptual processing is at the core of our approach:

Each node contains a floating-point activation value,  $\alpha$ , which represents its excitatory state, may affect its internal processing, and is in turn forwarded (potentially altered by the node's processing) to the node's afferent connections. In



**Fig. 1** A process node within a modality stream. Weighted activation travels both up from sensory events to concepts an actions (the afferent pathway), and—through simulation—back downstream (the efferent pathway)

a purely bottom-up framework, this activation represents the tiered analysis of sensory information towards action selection. Information flows up from sensory nodes, is distilled through feature, property, and concept nodes, and may activate one or more action nodes leading to motor activity.

A separate simulated activation value  $\sigma$  is also taken into account in the node's activation behavior and processing. This value represents simulated perception that does not directly stem from sensory data. Its source is internal to the cognitive system, and could originate, for example, from anticipation of a perceptual event, or intermodal priming.  $\sigma$  is added to the activation propagation when a node activates its afferent processing nodes. Also,  $\sigma + \alpha$  is used as a motor action trigger value in the action nodes. This dual mechanism allows us to model priming.

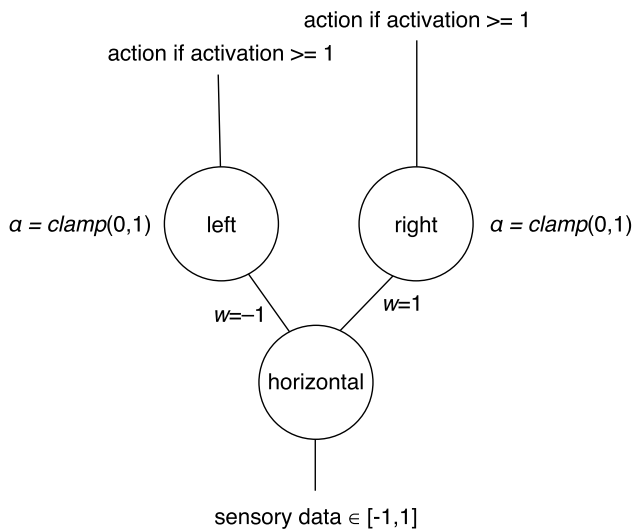
## 2.2 Priming

In humans, we observe the psychological phenomenon of “priming”, or the bias (often measured as a decrease in response time) towards a previously triggered sensory or recalled memory event. Such priming can occur through cross-modal activation, through previous activation, or from memory recall. Seen as a core element in fluent joint action, we can model priming through the efferent pathways in the modality streams:

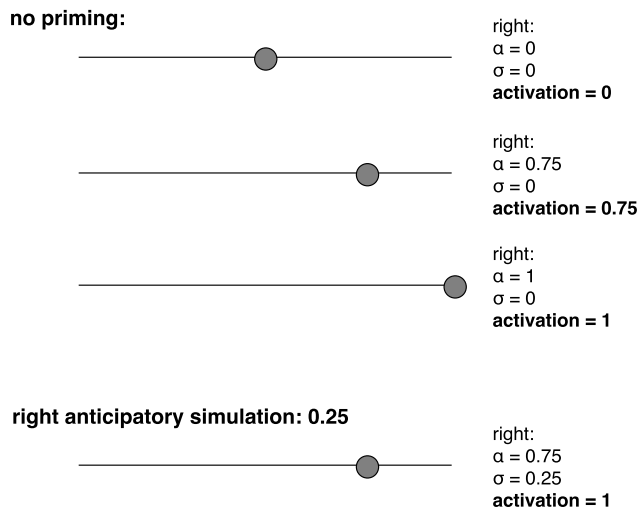
If a certain higher-level node  $n$  is activated through priming, the lower-level nodes that feed  $n$  are partially activated through the simulation value  $\sigma$  on the efferent pathway. As  $\sigma$  is added to the sensory-based activation  $\alpha$  in the lower-level nodes, this top-down activation inherently lowers the perceptual activation necessary for the activation of those lower-level nodes, decreasing the real-world sensory-based activation threshold for action triggering. The result of this is reduced response time for anticipated sensory events, and increasingly automatic motor behavior.

For example, let us assume a simple sensory activation stream which includes a sensor detecting the one-dimensional position  $x \in [-1, 1]$  of an object of interest (Fig. 2). This could be a goalkeeping agent playing a “Pong”-like game. The one-dimensional sensor feeds into two feature nodes, which correspond to the object being “left” or “right”. In this example, the activation  $\alpha \in [0, 1]$  of the “left” node would correspond to  $\max(-x, 0)$ , and the activation of “right” to  $\max(x, 0)$ . Therefore, the more “left” the object of interest is, the more the “left” feature node would be active, and vice versa for the “right” feature node. The two feature nodes feed, in turn, into a “left” action or a “right” action to be performed. This could be a force applied to the motor layer of the agent with the aim of intercepting the object of interest.

If, in this network, the robot was primed toward a “right” perception (for example, inter-modally through a vocal command, intra-modally through a related sensory or memory



**Fig. 2** A simple network illustrating the afferent and efferent pathways between process nodes. This network analyses the output from a one-dimensional sensor



**Fig. 3** Priming in the simple network shown in Fig. 2. Activation in intermediate nodes occurs earlier when sensory data matches anticipatory perceptual priming: Without priming (bottom-up processing), the robot will activate the action only when the sensory data is  $-1$  or  $1$ . When the “right” feature node is primed to the extent of  $0.25$ , less pronounced sensory data in the right direction can activate the “right” action

event, or—as we use it below—through anticipation), the efferent connection to the “right” feature node would partially activate by receiving a top-down simulation value  $\sigma$ , which would add to the sensory-based activation value. In that case, the partially activated feature node could become fully activated with a smaller value of  $x$ , resulting in an earlier appropriate action on the robot’s part. See Fig. 3 for an illustration of this example.

The result would be an agent that uses perceptual simulation to intercept faster-moving objects if it had an a-priori

bias, or prime, to the expected sensory data. This could model an agent that gets better at playing “Pong” through practice, or by using cross-modal perceptual information.

### 2.3 Practice subsystems

We have so far discussed the *effects* of simulated perception. This section discusses the *generation* of perceptual simulation for robot practice. In our framework, there are two top-down subsystems used to support practice within the proposed perceptual node architecture:

#### 2.3.1 Markov-chain Bayesian anticipatory simulation

The first subsystem is a Markov-chain Bayesian predictor, building a probabilistic map of node activation based on recurring activation sequences during practice. This system is in the spirit of the anticipatory system described in Hoffman and Breazeal (2007). This subsystem triggers high-level simulation (mainly in Action and Concept nodes), which—through modality stream’s efferent pathways—biases the activation of lower-level perceptual nodes.

If the subsequent sensory data supports these perceptual expectations, the robot’s reaction times are shortened as described above. In the case where the sensory data does not support the simulated perception, reaction time is longer and can, in some cases, lead to a short erroneous action, which is then corrected by the real-world sensory data. This subsystem corresponds to single-modal practice through repetition.

#### 2.3.2 Inter-modal Hebbian reinforcing

An additional mechanism of practice is that of *Hebbian reinforcement* on existing activation connections. While most node connections are fixed, some can be assigned to a connection reinforcement system, which will dynamically change the connection weights between the nodes. This system works according to the contingency principle introduced in Hebb (1949), reinforcing connections that co-occur frequently and consistently, and decreasing the weight of connections that are infrequent or inconsistent (the “fire together, wire together” principle).

More formally, let  $a_s$  denote the activation value of the upstream (or *start*) node of a connection governed by the connection reinforcement mechanism, and  $a_e$  the activation value of the downstream (or *end*) node of the same connection. Then, if  $a_s$  is greater than a threshold  $\epsilon$ , we compute  $\delta_{s \rightarrow e} = (a_s - \epsilon) \times (a_e - \beta)$ , for a second threshold  $\beta$ . This value  $\delta_{s \rightarrow e}$  is then added to the current connection weight.

Let  $\delta_s \equiv \max_e(\delta_{s \rightarrow e})$  and  $E_s \equiv \arg \max_e(\delta_{s \rightarrow e})$ .  $E_s$  is thus the downstream node of node  $s$  that applies for the highest reinforcement value at a given moment. In each update, we additionally decrease the connection weight of all connections  $s \rightarrow e$ , for  $e \neq E_s$ , by  $\delta_s$ .

This subsystem thus reinforces consistent coincidental activations, but inhibits competing reinforcements stemming from the same source node, leading to anticipated simulated perception of inter-modal perception nodes. This, again, triggers top-down biasing of lower-level perception nodes, shortening reaction times as described above. Please refer to Sect. 4.2.2 for a demonstration of the effect of repetition on connection weight.

### 3 Application

We have implemented an instantiation of the proposed architecture on a robotic system, which we subsequently used to evaluate our approach in a controlled human-subject study.

#### 3.1 Robotic platform

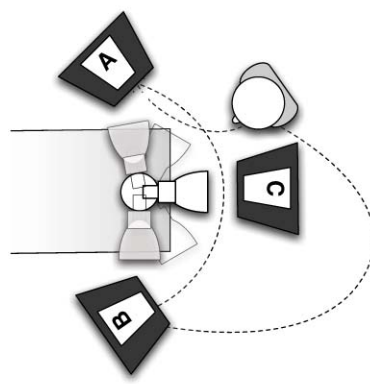
The robot employed in this evaluation was AUR, a robotic desk lamp, seen in Fig. 4(b). The lamp has a 5-degree-of-

freedom arm and a LED lamp which can illuminate in a range of the red-green-blue color space. AUR is stationary and mounted on top of a steel and wood workbench locating its base at approximately 90 cm above the floor. Its processing is done on a 2x Dual 2.66 GHz Intel processor machine located underneath the workbench.

The robot uses a Vicon motion capture system to identify and track the location and orientation of the human’s right hand at a frequency of 10 times per second. This was made possible by a special glove with retroreflective markers on it, worn on the human’s right hand.

The system also takes input from Sphinx-4, an open-source speech recognition system created by the Sphinx group at Carnegie Mellon University, in collaboration with Sun Microsystems Laboratories, Mitsubishi Electric Research Labs, and Hewlett Packard (Walker et al. 2004). The commands recognized by the system in this task were: “Come”, “Come Here”, “Go”, “Red”, “Blue”, “Green”, and “Off”.

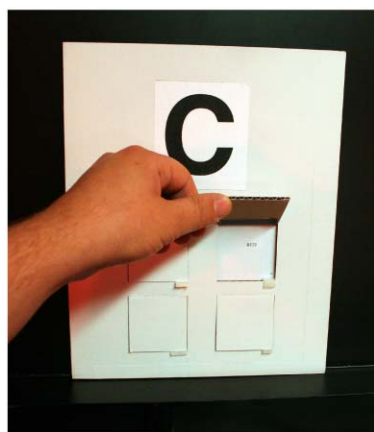
**Fig. 4** (a) Diagram, and (b) photograph of the collaborative lighting task workspace. (c) Carboard at each task location. (d) Sample experimental sequence



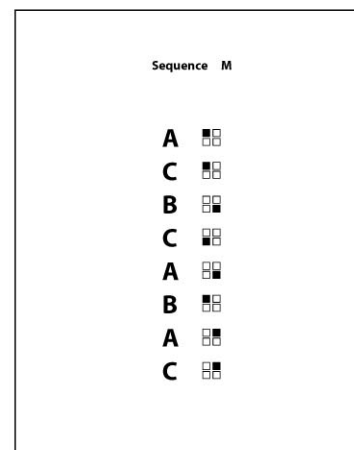
(a)



(b)



(c)



(d) Sequence M

### 3.2 Task description

In the human-robot collaboration used in our studies the human operates in a workspace as depicted in Fig. 4(a) and (b). The robot could direct its lampshade to different locations around its own axis, and change the color of the light beam. When asked to “Go”, “Come”, or “Come here” the robot would move to the location of the person’s hand, assuming the hand was relatively static. Additionally, the color changed in response to speech commands to one of three colors: blue, red, and green.

The workspace contained three locations (A, B, C). At each location there was a white cardboard square labeled with the location letter, and including four doors (Fig. 4(c)). Each door, when lifted, revealed the name of a color written underneath. The task was to complete a sequence of 8 actions, which was described in diagrammatical form on a sequence sheet as shown in Fig. 4(d). This sequence was to be repeated 10 times, as quickly as possible.

Each action in the sequence specifies: a general location A, B, or C, and an indication of which of the four doors to open. The action is completed when the lamp shines the specified color of light at that location. This would result in the sound of a buzzer, indicating the person should move to the next action in the sequence. A different buzzer was sounded when a whole sequence was completed. Neither the human, nor the robot, know the order of actions in the task sequences, or the names of the colors hidden behind the doors, at the beginning of the task.

## 4 Cognitive network

To solve the task described in this paper, we designed a cognitive network along the principles introduced in Sect. 2.

### 4.1 Modality streams

The network is made up of three modality streams: a visual, an auditory, and a proprioceptive stream. The action network includes five actions, three color changing nodes, a color “off” node, and a “goto position” node.

#### 4.1.1 Visual modality

The visual modality stream’s sensory input stems from the Vicon motion capture system, indicating the hand position in the workspace. Two feature areas are downstream from that Vicon sensory node: in one, four workspace segmentation feature nodes activate proportionally to the proximity of the hand to each of the four corners of the workspace. In the other, a speed node detects the speed as a single-frame position derivative of the hand position. Downstream from the

speed node, an inhibitory connection feeds the “Stillness” feature node, which simply detects the inverse of the speed node clamped between 0 and 1. The stillness node feeds into a property node indicating whether the hand has been still for some consecutive period of time. Three concept nodes represent the location of the hand near one of the task targets, A, B, or C. Note that the robot does not know the location of the targets ahead of time, but infers them from combination of stillness and hand position feature nodes.

The target concept nodes are connected to the “goto position” action node, as is a conjunction node combining the “stillness settled” and the “Go” speech feature node. A combination of a hand position near the target, an aggregate stillness of the hand, and a “Go” command will trigger this action, leading the robot to move its beam towards the appropriate position.

#### 4.1.2 Auditory modality

The auditory modality stream uses input from the Sphinx speech recognition system, parsing the auditory input into speech tokens. This sensory node feeds five speech feature nodes, which respond to the detection of a particular speech token in the auditory stream. Four of the speech feature nodes have afferent connections to the four light change action nodes, the fifth leads, through, a conjunction node as described above, to the “goto position” action node.

#### 4.1.3 Proprioceptive modality

Finally, the proprioceptive modality stream is based on the joint positions of the robot, thus representing the robot’s sense of physical self. This simple modality stream has a direction feature node, which calculates the lamp head direction from the joint positions. This node, in turn, feeds into left, right, and center property nodes, which classify the overall orientation of the robot.

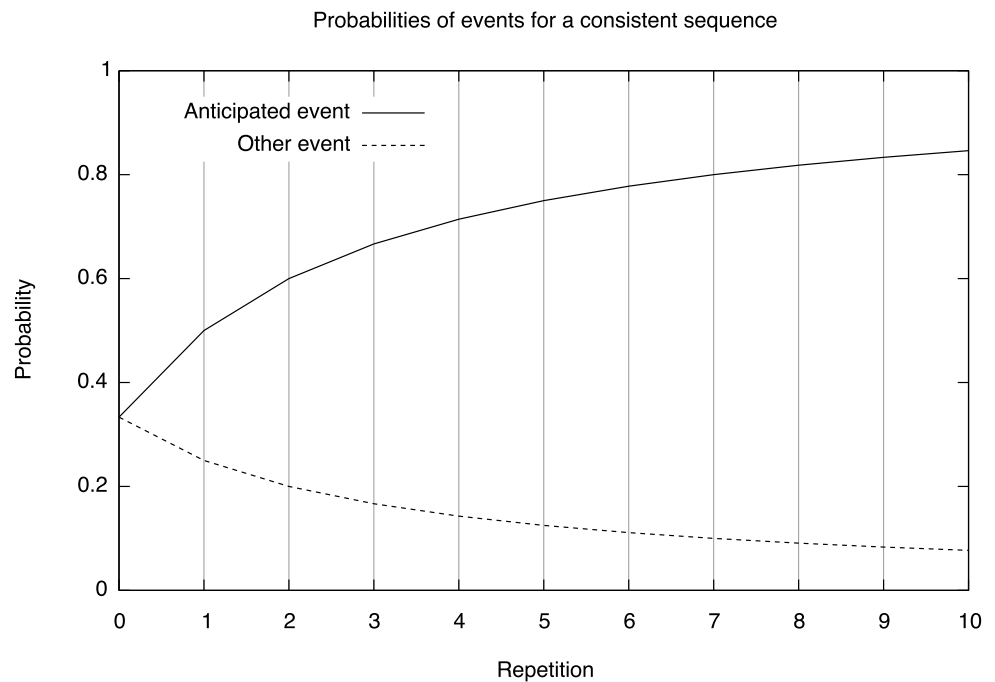
### 4.2 Practice subsystems

The implementation of the practice subsystems for this task is as follows:

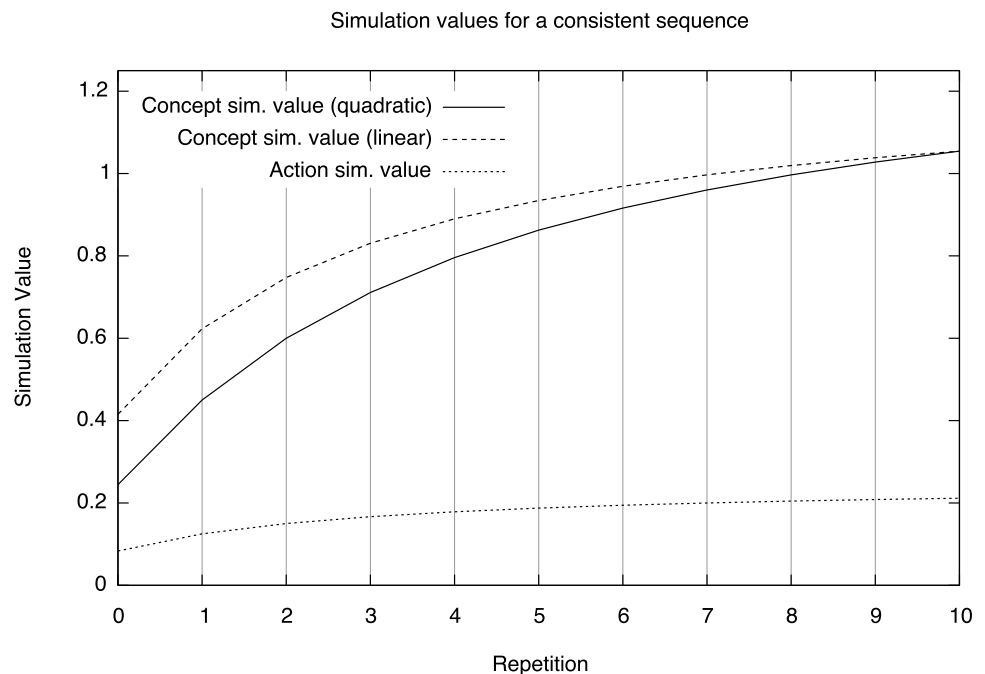
#### 4.2.1 Markov-chain Bayesian predictor

Using a 3-step sequence history Markov model, the learner estimates the probability of the appropriate target board, and the “Go” action being expected. The probability of a certain concept to be triggered next, denoted  $p$ , is translated into a simulation value  $\sigma_c = (p + \gamma) \times p$ . In our experiments, we used  $\gamma = 0.4$ . For the action simulation factor, we used a linear coefficient,  $\sigma_a = p \times \gamma$ . In the experiments described below,  $\gamma = 0.25$ .

**Fig. 5** Probability for an anticipated event and a not anticipated event in case of a consistent human, based on the Markov-chain Bayesian predictor



**Fig. 6** Simulation values for anticipated concept and action with a consistent human teammate, based on the Markov-chain Bayesian predictor. Compare the quadratic concept simulation value used in our experiments with a linear derived activation converging on the same number



This simulated value then moves down the efferent modality stream, biasing visual perceptual nodes. As a result, feature nodes simulate to the extent that they are correlated with the appropriate hand-target concept. Thus an increasing distance between the hand position and the correct target is adequate to trigger the appropriate response, and robot reaction time is decreased.

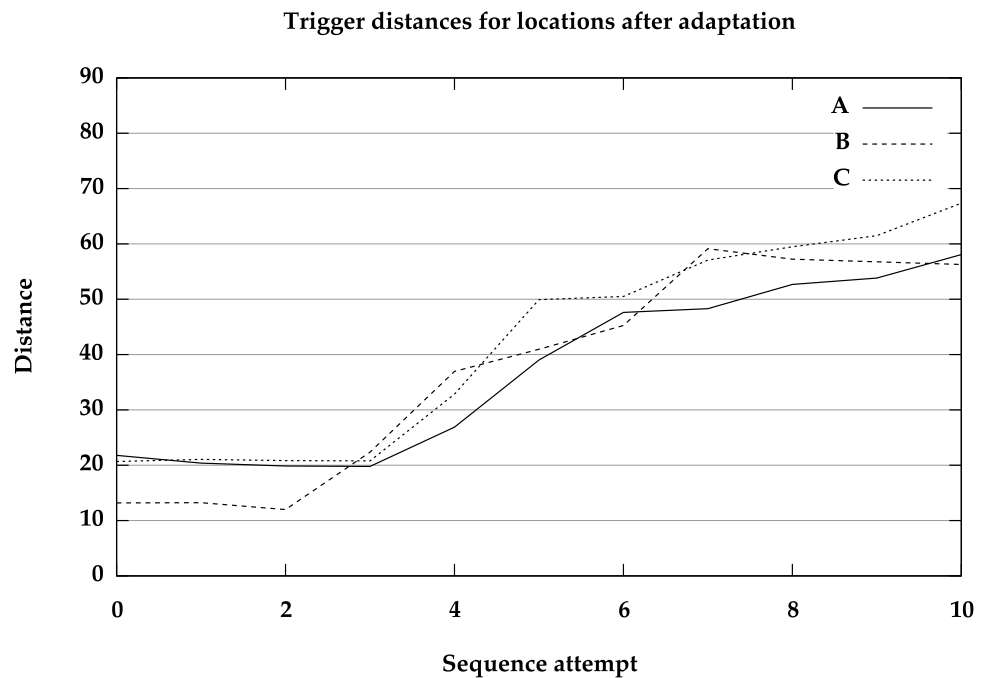
Figure 5 shows the probabilities of anticipated concepts for a consistent human teammate. The graph shows these

probabilities at the same step for each attempt for a given expected concept, and for a not-anticipated concept.

Figure 6 shows the simulation value associated with the probabilities shown in Fig. 5. Compare the quadratic concept simulation value with a linear derived activation converging on the same number.

Figure 7 shows the change in trigger area for each of the areas using a simple A-C-B sequence with ten iterations. The longer the practice session, the more “primed” the per-

**Fig. 7** Effect of emulator on trigger distances for each of the three locations using a simple A-C-B sequence



ceptual stream is, and the further away the hand position can be to result in a concept and subsequent action activation. If, at the beginning of a session, the human’s hand needs to be as close as 20 cm from the target for the robot to move, after 8 consistent trials, the robot responds to a hand position as far as 60 cm, granted that this move is towards an anticipated target location.

As described in Sect. 2.3.1, we found that if the human moves in the wrong direction for a certain next step, in many cases the robot is triggered to move briefly in the correct direction before following the human’s command. This often results in a joint matched movement to one and then to the other direction performed by both the human and the robot, a “double-take” of sorts. We believe that such an embodied mirroring behavior could play a role adding to the team’s sense of bond, as well as to the human’s perception of the robot as similar to themselves. We further explore this notion in our experiments below.

#### 4.2.2 Hebbian reinforcement

In addition, we used inter-modal simulation, affected by the Hebbian process described above, between the robot’s proprioceptive property nodes, which sense the robot’s joint configuration (“Left”, “Right”, and “Center”), and the auditory feature nodes, as described in Sect. 2.3.2. Thus, certain physical configurations of the robot lead to the simulation of a certain word in the auditory stream, resulting in the perceptual simulation of that speech segment when the robot reaches a certain position. If there is a consistent correlation between position and color, the robot will increasingly

trigger the appropriate color without an explicit human command.

The effects of connection reinforcement on inter-modal simulation can be seen in Fig. 8. The first graph shows the weights between the Center Property Node and each of the color speech feature nodes with a mostly consistent mapping between the two modality instances.

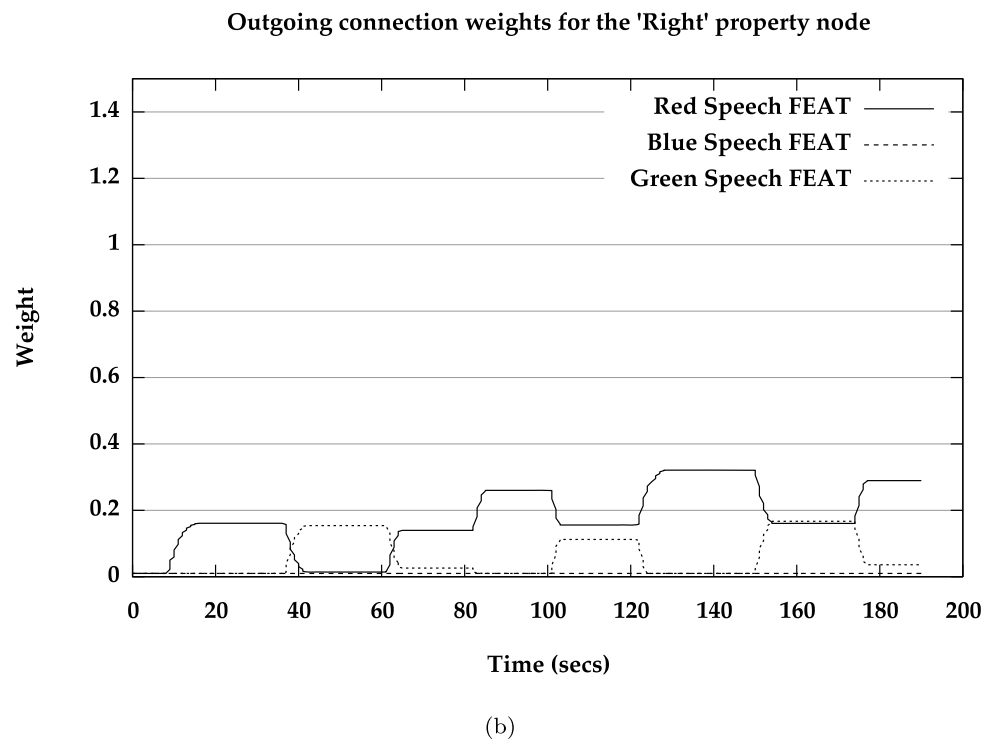
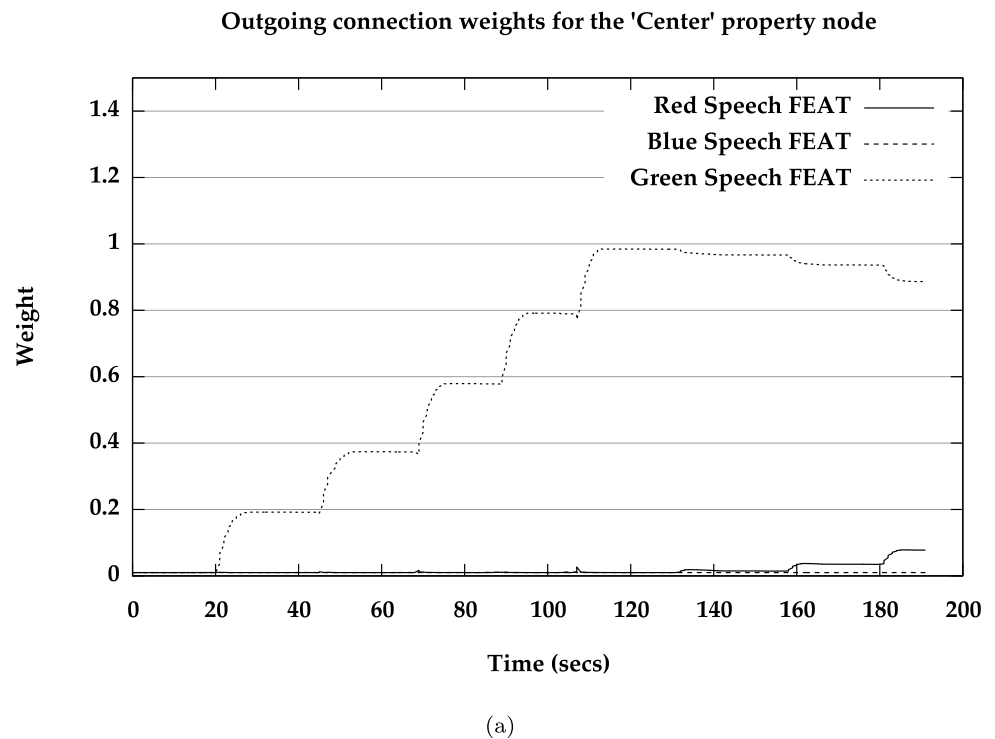
The second graph shows the effects of mostly alternating contingency between the proprioceptive node and the speech feature nodes. Across the experimental sequence, these reinforcements cancel each other out and result in a low simulation factor between these nodes.

## 5 Experimental design

To evaluate the validity of our approach to human-robot teamwork, we conducted a between-group controlled experiment with two conditions. The control (or REACTIVE) condition corresponds to the baseline condition in which no anticipatory simulation or cross-modal reinforcement occurred. The remainder of the system, i.e. the perceptual network and all activation streams and thresholds, were identically retained. In the second (FLUENCY) condition, the simulation subsystems were active with fixed parameters.

To control for instruction bias, neither group was told whether the robot will adapt to their behavior. All participants were allowed to practice with the system before beginning the experiment. The full instructions of the experiment can be found in Hoffman (2007).

**Fig. 8** Hebbian reinforcement between the proprioceptive and auditory modality with (a) predominantly consistent contingency and (b) inconsistent contingency



The experiment included two sequences, or “patterns”. Pattern *A* was associated with a setup in which there were different colors under the different doors. In Pattern *B* there was a one-to-one mapping between location and color, i.e., the same color was hidden under all four doors in a single

location. For example, all doors in location *B* hid the color “Blue”, and all doors in location *A* hid the color “Green”. Therefore both the human’s memory and the robot’s inter-modal reinforcement process could more easily learn the correct association between spatial location and color.

We recruited 38 subject from the campus and area communities, through email solicitation. They were arbitrarily designated to one of the two experimental conditions, and for each subject, they were arbitrarily assigned the order of sequences between Pattern *A* first, and Pattern *B* first. At the last day of the experiment we experienced an unrecoverable hardware failure, forcing us to release the last 5 subjects. We thus remained with 33 subjects (17 male), 15 in the REACTIVE condition, and 18 in the FLUENCY condition. Two additional subjects in the FLUENCY condition experienced a mechanical failure. This was resolved in a short period of time, and the subjects continued the experiment. The failure affected the amount of data we were able to obtain from these subjects, a fact we addressed as described below.

This experimental protocol was reviewed and approved by the institutional review board of the Massachusetts Institute of Technology.

In sections below we will use the following terminology:

**Turn** is the time and actions occurring between two consecutive turn buzzers. These include a single event of correctly shining the right light onto the right board.

**Sequence** is a set of eight turns. There are ten *attempts* at a sequence.

**Round** is a set of ten attempts at a sequence. There are two rounds in each experiment. The *first round* is the round performed first; similarly for the phrase *second round*. Rounds can also be identified by *patterns*. In this case we will refer to *Pattern A* and *Pattern B*. Note that for different subjects the order of patterns, i.e. the mapping between patterns and round numbers, is different.

**Task** is a set of two rounds, using both patterns.

## 6 Results

The behavioral measures recorded in this experiment were elicited from the log files generated by the experiment software. The software running the robot logged a number of events, including the human's speech commands, the robot's action selection, the human's hand position, and the experimenter's buzzer times.

To account for a number of mechanical failures as mentioned above, as well as for mistakenly recorded turn and sequence buzzer events, the data has been automatically cleaned up, by eliminating the following sequences: (a) any sequence attempt that does not contain 7–9 turns was eliminated; (b) any sequence attempt that lasted for less than 25 seconds or more than 180 seconds was eliminated.

As a result, two subjects' data included only 8 sequences in one of the rounds, one subject's data remained with 9

sequences in both rounds, and two subjects' data remained with 9 sequences in one of their two rounds.

We have included the valid data from these subjects in our analyses, except in Hypothesis H1 below. In H1, as well as in the graphs depicting sequence-by-sequence progress on the recorded behavioral measures, we included only data from trials containing 10 valid sequences.

### 6.1 Team performance

Our first set of hypotheses were concerned with the performance of the human-robot team. We hypothesized the following metrics to be significantly lower in the FLUENCY condition compared to the REACTIVE condition:

**H1** The overall task completion time (both rounds).

**H2** Mean sequence attempt time.

**H3** Mean sequence attempt time (second half).

**H4** Best sequence attempt time.

**H5** Improvement (ratio between last and first attempt).<sup>1</sup>

Using a T-test with independent samples, we find a significant difference between the two conditions in all five hypotheses. All values are in seconds, except in H5, which is a fraction.

**H1** Total task time: REACTIVE:  $1401.66 \pm 162.90$ , FLUENCY:  $1196.26 \pm 226.83$ ;  $t(24) = 2.609$ ,  $p < 0.05$ .

**H2** Mean sequence time: REACTIVE:  $141.38 \pm 17.38$ , FLUENCY:  $116.21 \pm 22.67$ ;  $t(30) = 3.487$ ,  $p < 0.01$ .

**H3** Mean sequence time (2nd half): REACTIVE:  $131.04 \pm 17.76$ , FLUENCY:  $97.93 \pm 22.75$ ;  $t(30) = 4.544$ ,  $p < 0.001$ .

**H4** Best sequence time: REACTIVE:  $117.93 \pm 16.11$ , FLUENCY:  $83.87 \pm 18.81$ ;  $t(30) = 5.461$ ,  $p < 0.001$ .

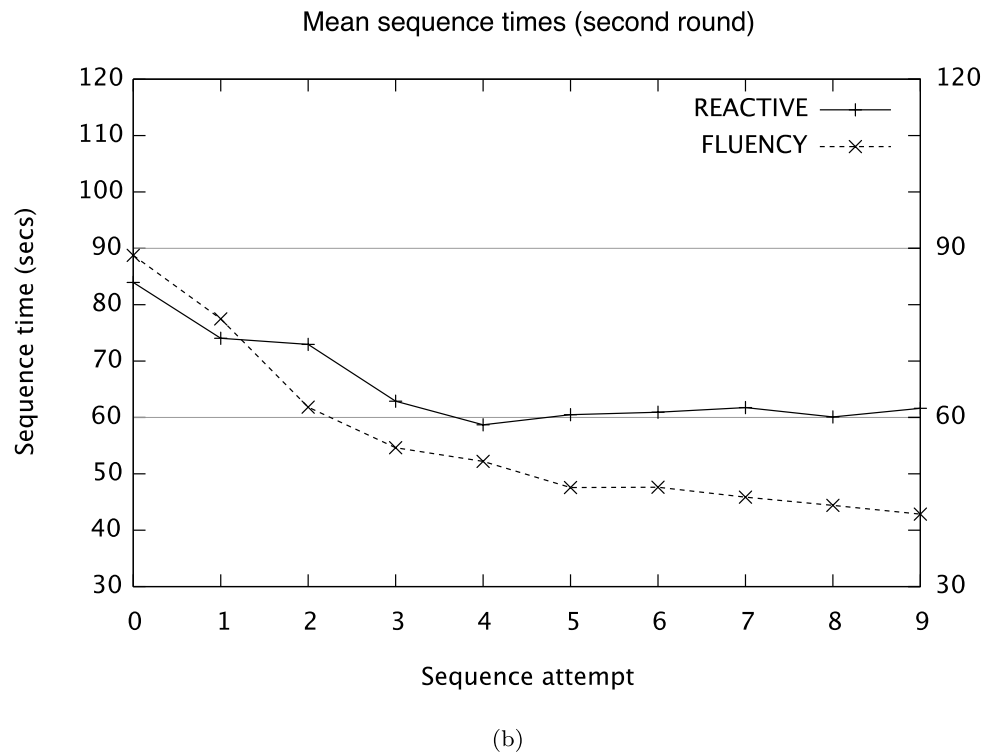
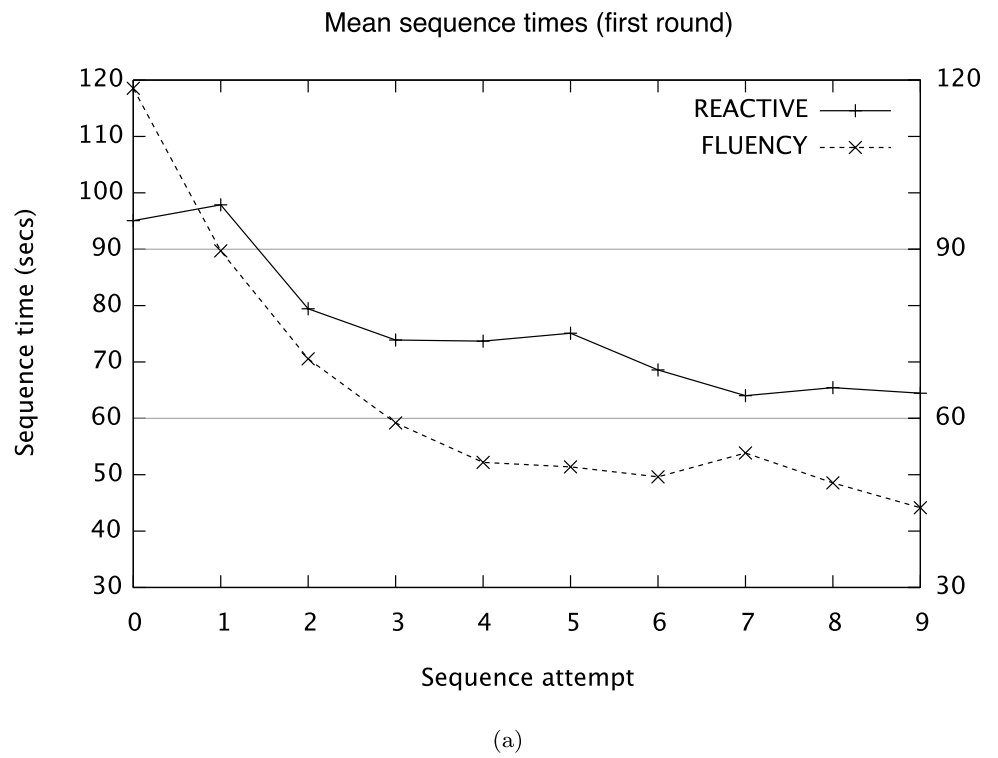
**H5** Sequence time improvement: REACTIVE:  $0.76 \pm 0.18$ , FLUENCY:  $0.50 \pm 0.15$ ;  $t(30) = 4.718$ ,  $p < 0.001$ .

All of our hypotheses were confirmed, demonstrating a significant improvement in team performance under the FLUENCY condition. Note also, that examining the second half of each task round leads to an increase in difference between the two conditions, and an increase in significance.

Figure 9 shows the average sequence attempt time for both conditions, split by round. The notion of initial practice runs is evident in this figure, as the second round starts at a lower time than the first round. In both cases, the FLU-

<sup>1</sup>We use only data from the second round under the assumption that, at this point, the subject is familiar with the task structure and robot, and we thus measure only the team's improvement and not the initial practice needed by the subject.

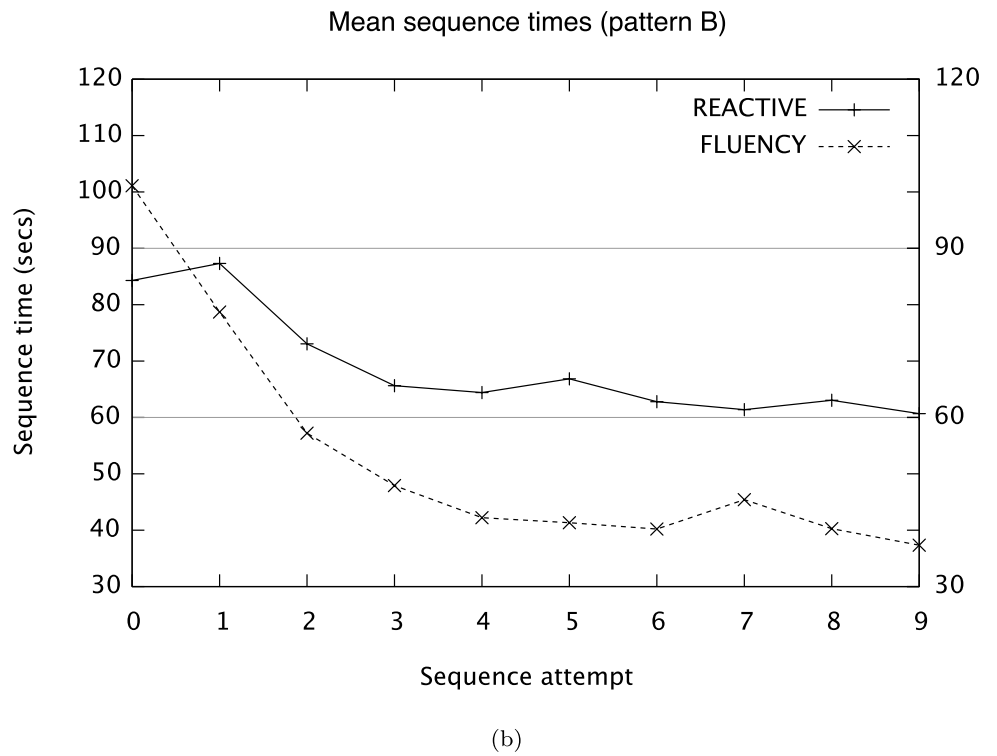
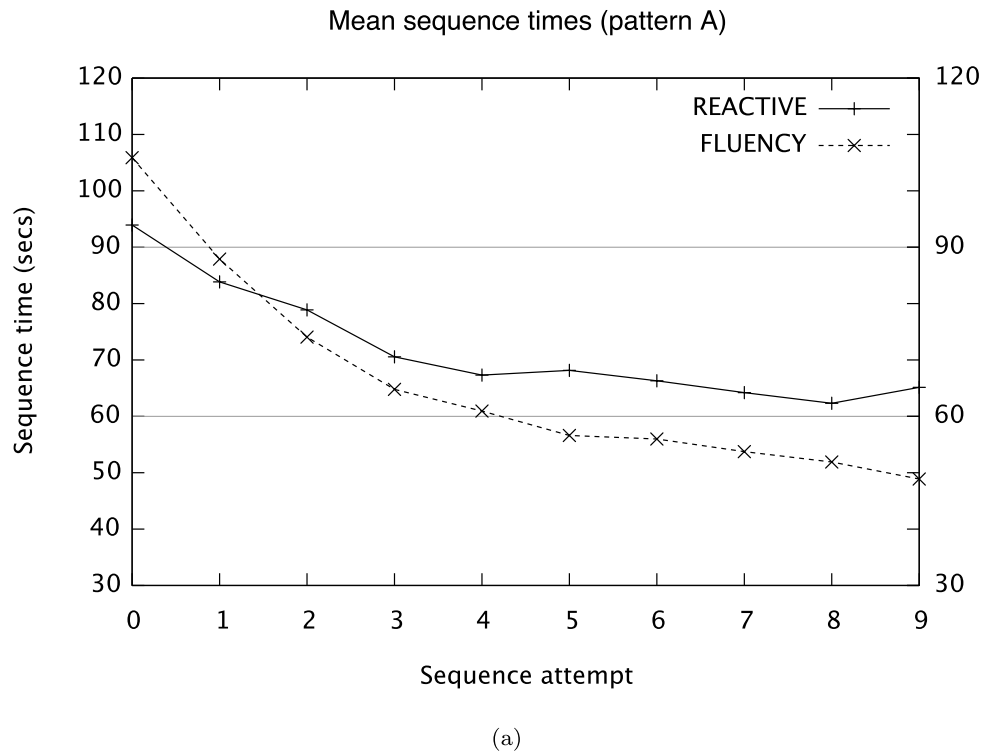
**Fig. 9** Mean sequence times—per round—over two ten-attempt practice sessions, comparing the REACTIVE and FLUENCY conditions



ENCY condition converges at slightly over 40 seconds, while the REACTIVE condition maintains an average over 60 seconds.

Figure 10 shows data for sequence pattern A and B, respectively. Since the robot and the human “learn” the color sequences more easily in pattern B, we see a more dra-

**Fig. 10** Mean sequence times—per pattern—over two ten-attempt practice sessions, comparing the REACTIVE and FLUENCY conditions



matic improvement in the FLUENCY condition, converging on a below-40 second score in the final sequence attempt.

### 6.2 Fluency metrics

The second set of hypotheses tested in this experiment relate to the fluency of the team. In Hoffman and Breazeal (2007)

we found that in some cases, even when there is no improvement in the efficiency of the team’s performance, subjects perceive a significant difference in the fluency of the team. In that work, we found two fluency metrics to be correlated to the human’s sense of fluency. In the experiment presented in this paper, we elicited the same metrics from the log files recorded by the experiment software:

**IDLE.** Human idle time. The percentage of time within each task round in which the human hand was stationary or moved very little. This metric is elicited from the human’s hand position. A low-passed sensor with hysteresis is triggered every time the frame-by-frame distance of the hand position crosses a certain threshold.

**DELAY.** Robot functional delay. The time that passed from the beginning of a turn to the onset of the robot’s movement.

We tested the hypotheses that the following metrics are significantly lower in the **FLUENCY** condition compared to the **REACTIVE** condition:

**H6** Mean human idle time.

**H7** Mean robot functional delay.

**H8** Mean robot functional delay (second half).

We found significant difference between the two conditions on these three hypotheses, as well. Using a T-test for independent samples, H6 is a ratio, H7 and H8 are in seconds:

**H6** Human idle time: **REACTIVE**:  $0.46 \pm 0.08$ , **FLUENCY**:  $0.364 \pm 0.09$ ;  $t(30) = 3.001$ ,  $p < 0.01$ .

**H7** Robot functional delay: **REACTIVE**:  $4.81 \pm 9.91$ , **FLUENCY**:  $3.66 \pm 15.72$ ;  $t(30) = 2.434$ ,  $p < 0.05$ .

**H8** Robot functional delay (2nd half): **REACTIVE**:  $4.07 \pm 10.97$ , **FLUENCY**:  $1.48 \pm 6.14$ ;  $t(30) = 3.487$ ,  $p < 0.001$ .

All of our hypotheses were confirmed, demonstrating a significant improvement in both fluency metrics under the **FLUENCY** condition. Again, examining the second half of each task round (Hypothesis H8) shows an increase in difference between the two conditions, and an increase in significance.

The two evaluated fluency metrics are interrelated: since subjects have been instructed to complete the task as quickly as possible, a lower functional delay on the robot’s part would also result in less idle time on the human’s part—specifically the time that the human waits for the robot to start moving. However, in previous work, we have shown that these phenomena can occur even when the overall pace is not increased, and separate from each other.

Based on the questionnaire responses discussed below, we are led to believe that part of the human’s improvement in promptness can be attributed to the increased “performance pressure” exerted on the human team-mate, when the

robot shows a noticeable improvement in reaction time. As shown below, this can have both positive and negative effects on the human’s perception of the robot and of themselves.

Figures 11 and 12 shows the average change in robot functional delay time for both conditions, split by trial. It can be seen, in Fig. 11(b), that in the **REACTIVE** condition, the human teammate can do little to improve the robot’s delay, after the initial practice period.

The efficiency results above have also been reported in Hoffman and Breazeal (2008).

### 6.3 Relative contribution of human and robot

As both the human and the robotic team members undergo a learning curve of adapting to the collaborative task, it is interesting to estimate the relative contribution of each team member to the improvement of the team, comparing the learning rate of the human and the robot.

We estimated this measure as follows: For each sequence attempt, we compare two of the above-mentioned metrics to the value of the same metric in the first attempt in a given round. Since the first round included a few practice attempts, we only estimate this measure on the second round for each subject.

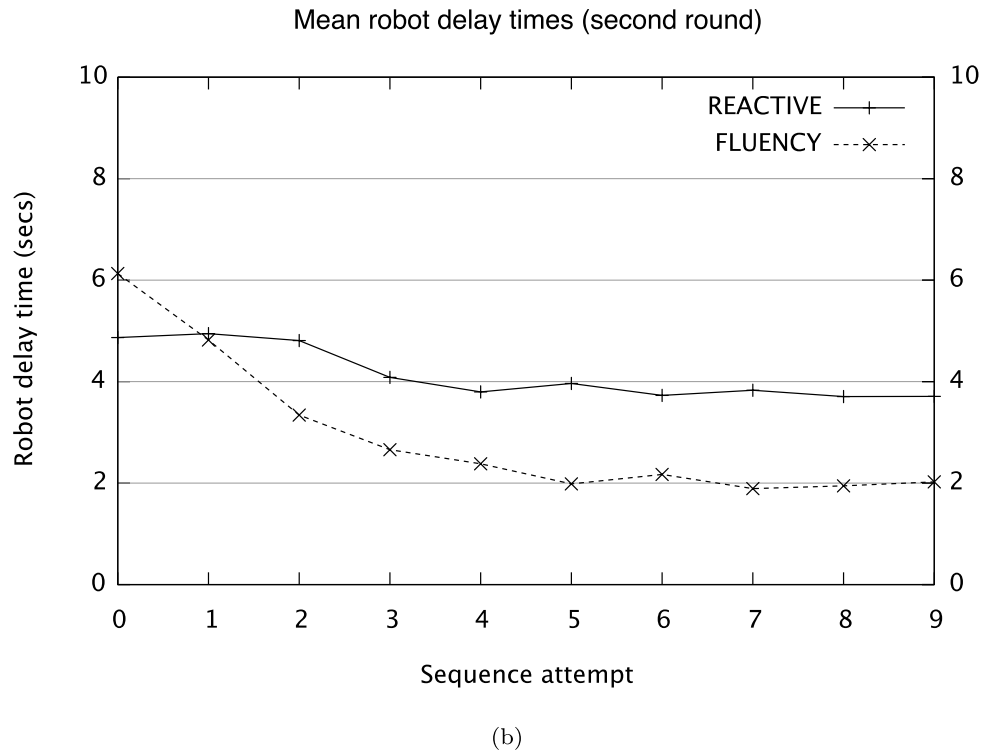
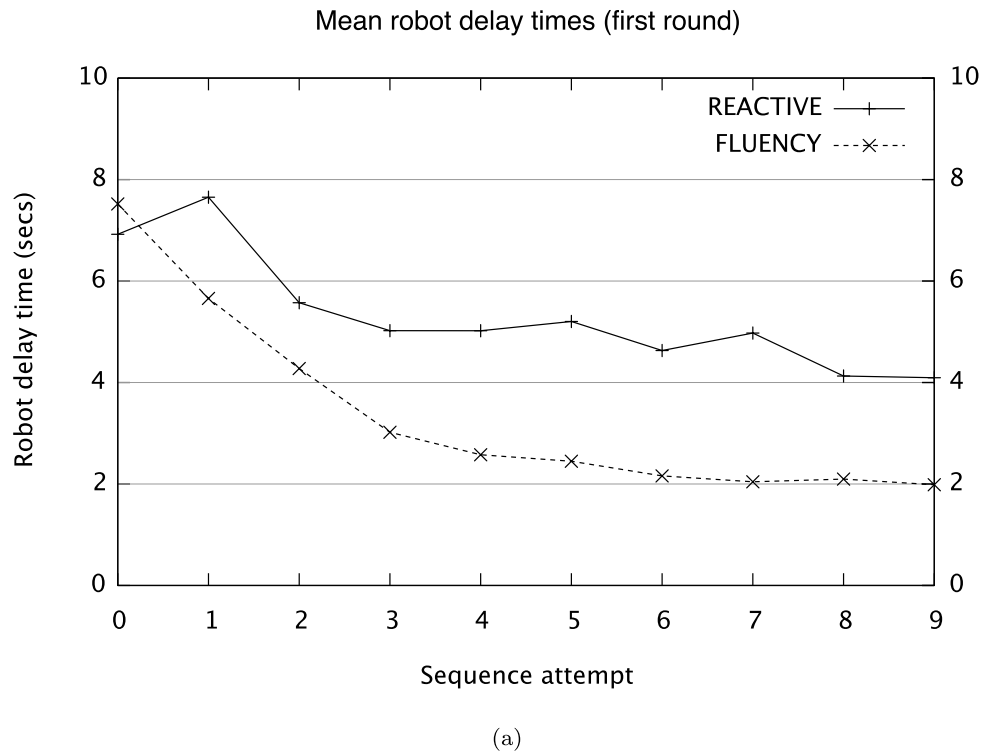
As the robot does not adapt or learn in the **REACTIVE** condition, we consider the improvement of the team in that group to be solely on behalf of the human. We call this “the human contribution” to the team’s improvement. Subtracting the human contribution function from the improvement of the team in the **FLUENCY** condition, we obtain “the robot contribution” to the team’s improvement.

Figure 13(a) shows the relative contribution of the team members on the improvement in sequence time. We find that the rate of adaptation on the robot’s part roughly matches that of the human, both contributing to about 20% of the reduction in sequence time over the course of a ten-attempt experimental round. We postulate that this phenomenon may contribute to an increased sense of partnership and “like-me” perception in human-robot teams.

So far there has been virtually no discussion in the literature about the effects of matching learning rates in tasks where both human and robot improve on a joint task. We find these initial results to provide a promising foundation for future research in this area.

In Fig. 13(b) we show the robot’s contribution to the robot’s functional delay (a measure that has been shown to be related to team fluency). Again, we see a similar adaptation curve, but on this metric the robot’s contribution converges on roughly twice that of the human, contributing to a circa 40% improvement compared to the human’s circa 20%.

**Fig. 11** Mean functional delay times—per round—over two ten-attempt practice sessions, comparing the REACTIVE and FLUENCY conditions

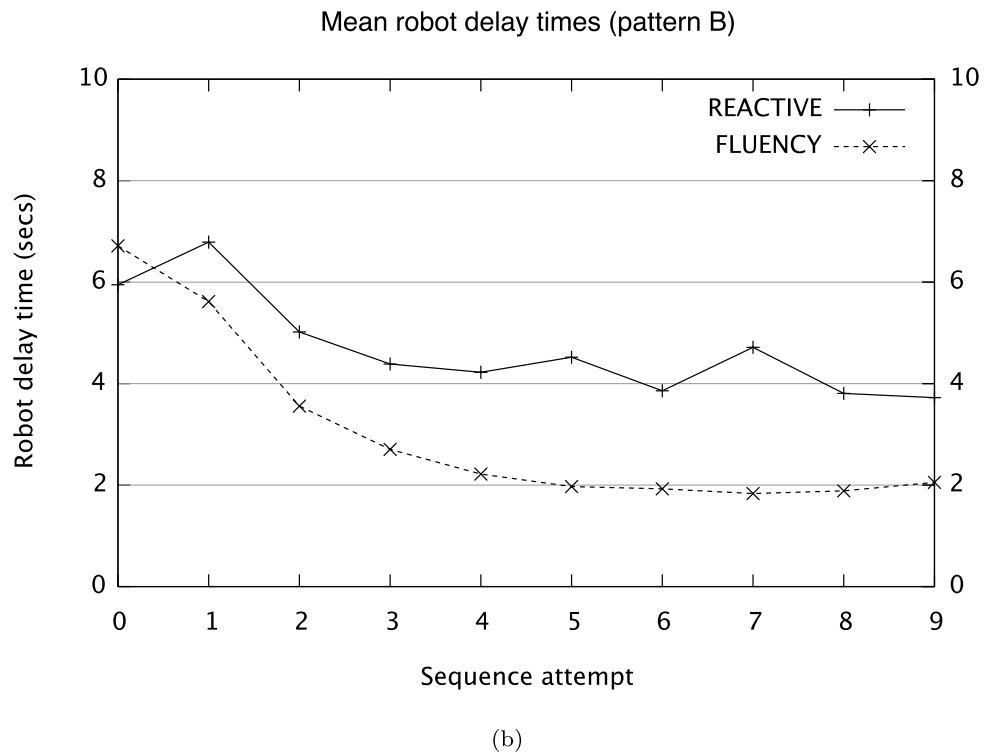
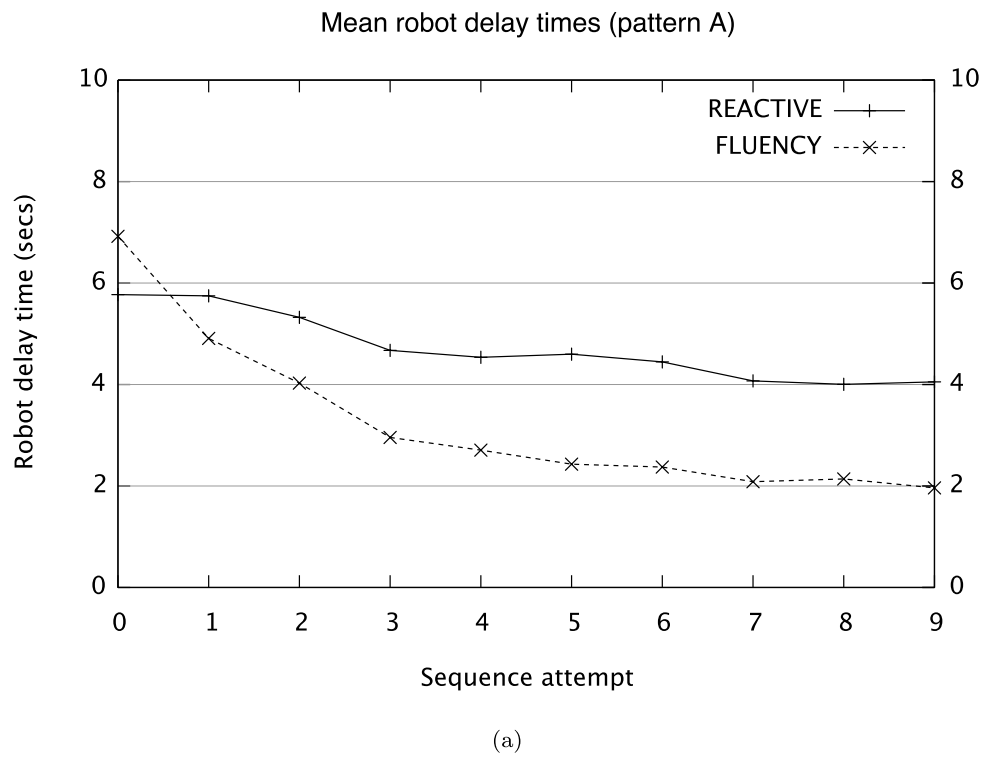


#### 6.4 Self-report questionnaire

In addition to the behavioral metrics we have administered a self-report questionnaire including 41 questions. These

questions were aimed to evaluate the human teammates' reaction to the robot with and without perceptual simulation. 38 questions asked the subjects to rank agreement with a sentence on a 7-point Likert scale from "Strongly Dis-

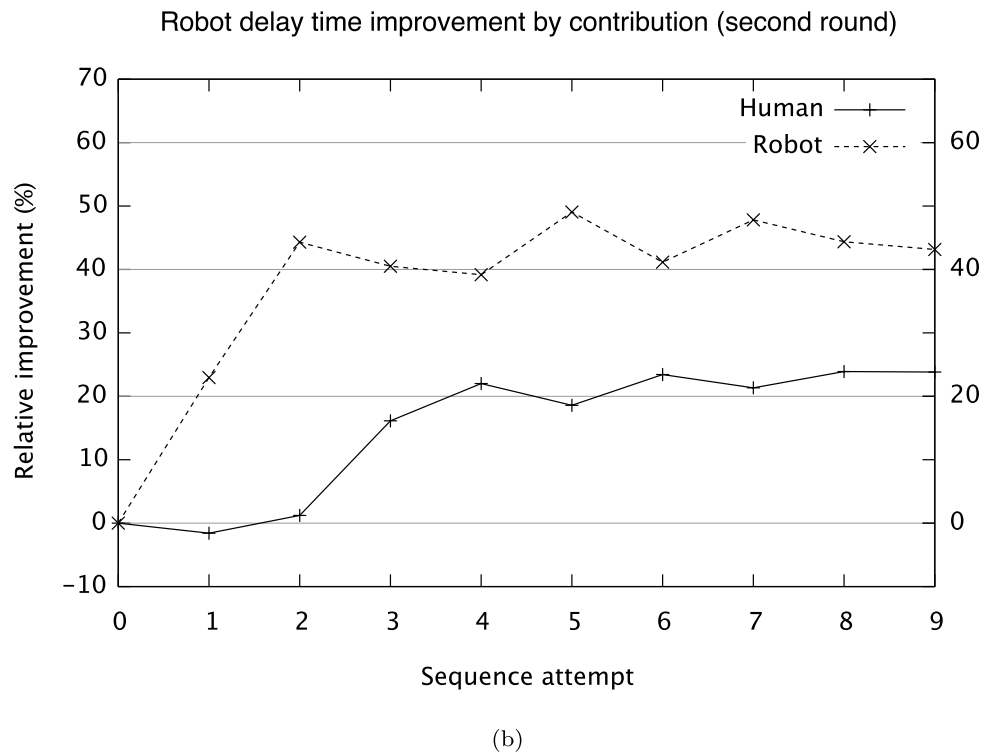
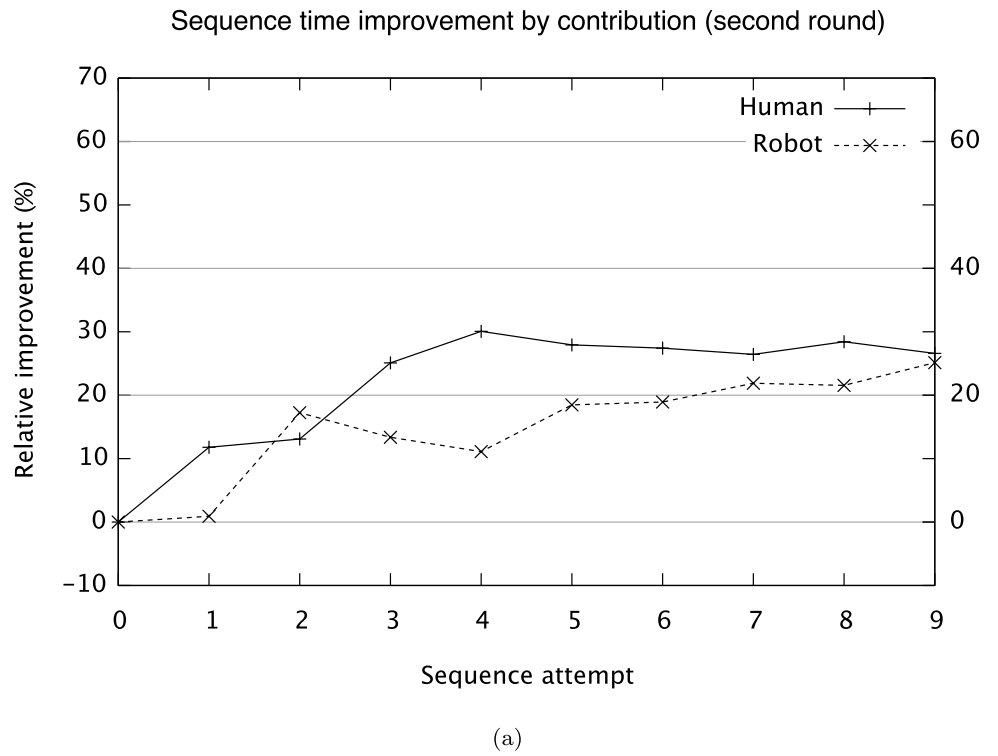
**Fig. 12** Mean functional delay times—per pattern—over two ten-attempt practice sessions, comparing the REACTIVE and FLUENCY conditions



agree” (1) to “Strongly agree” (7). Three questions were open ended responses. We have compounded the 38 scaled questions into eight compound scales we propose to be valu-

able to evaluate human-robot teamwork, and have verified the reliability of these scales within our subject population using Cronbach’s alpha measure.

**Fig. 13** Relative contribution of the team members on (a) sequence time, and (b) robot delay



- FLUENCY. The sense of fluency in the teamwork;
- IMPROVE. The team’s improvement over time;
- ROB-CONTRIB. The robot’s contribution;
- ROB-TRUST. The human’s trust in the robot;

- ROB-CHAR. The robot’s positive character traits;
- WAI-BOND. The Working Alliance bond subscale;
- WAI-GOAL. The Working Alliance goal subscale;
- WAI. The overall Working Alliance.

**Table 1** Survey questionnaire results metrics. Values are mean  $\pm$  s.d. on a 7-point Likert scale

Metric	REACTIVE	FLUENCY	t(31)	
FLUENCY	4.98 $\pm$ 0.96	5.93 $\pm$ 0.98	2.80	**
IMPROVE	5.16 $\pm$ 0.96	6.17 $\pm$ 1.09	2.80	**
ROB-CONTRIB	2.85 $\pm$ 1.11	4.00 $\pm$ 1.32	2.69	*
ROB-TRUST	4.90 $\pm$ 1.25	5.42 $\pm$ 1.28	1.17	
ROB-CHAR	4.80 $\pm$ 1.32	5.41 $\pm$ 1.17	1.25	
WAI-BOND	4.17 $\pm$ 1.05	4.37 $\pm$ 1.03	0.53	
WAI-GOAL	3.64 $\pm$ 1.47	4.70 $\pm$ 1.31	2.19	*
WAI	4.14 $\pm$ 0.94	4.54 $\pm$ 0.94	1.22	
ROB-FLUENCY	4.73 $\pm$ 1.22	6.11 $\pm$ 1.18	3.28	**
HUM-COMMIT	6.40 $\pm$ 0.74	5.83 $\pm$ 1.10	1.70	
ROB-ADAPT	3.47 $\pm$ 1.46	5.94 $\pm$ 1.06	5.66	***

The last three scales were adapted from the Working Alliance Inventory (Horvath and Greenberg 1989), a standard instrument evaluating clinician-patient relationship, to fit a human-robot joint task. We did not include a task subscale as the questions in the original WAI task subscale were very specific to a clinician-patient scenario.

In addition we evaluate the following individual questions, which were not compounded into scales:

- ROB-FLUENCY. The robot's contribution to the fluency;
- HUM-COMMIT. The human's commitment to the task;
- ROB-ADAPT. The robot's adaptation to the human.

We hypothesized there to be a significant difference in these metrics between the two conditions, and specifically that these metrics be higher for the FLUENCY condition.

Table 1 shows the results for the questionnaire hypotheses, and reveals significant differences between subjects in the two experimental conditions with regard to the fluency scales in the questionnaire. Both the FLUENCY and the ROB-FLUENCY measures are significantly different at  $p < 0.01$ . Additionally, subjects in the FLUENCY condition rated the robot's contribution to the team significantly higher than subjects in the REACTIVE condition, as well as the team's overall improvement. This supports our hypothesis that the proposed architecture contributes to the quality of fluency and collaboration in human-robot teams.

While these task-related scales differ significantly, we were not able to show a significant difference in the robot's compound positive character traits (intelligence, trustworthiness, and commitment), in the trust the human put in the robot, or in the human's commitment to the task—which was incidentally higher for the REACTIVE condition, if not significantly so. We believe that this is in part due to the low expectation people have of robots, which caused the evaluation of the REACTIVE robot to be high as a response to the robot's generally reliable functioning.

This hypothesis could be evaluated in a separate within-subject experiment comparing the two robot architectures. Also note that while the overall robot's character was not rated significantly different between the two conditions, the robot's intelligence was (REACTIVE:  $4.2 \pm 1.7$ , FLUENCY:  $5.33 \pm 1.08$ ,  $t(31) = -2.32$ ,  $p < 0.05$ ).

On the WAI scale, the goal subscale was significantly different between the two conditions, while the bond subscale—as well as the overall WAI score—were not. One possible explanation for that phenomenon could be that it takes longer than the experiment's duration to form a bond, whereas the mutual agreement on goals can be established in a shorter time span.

#### 6.4.1 Open-ended responses

The qualitative response of subjects in the open-ended responses of subjects in the FLUENCY condition was more favorable than that of subjects in the REACTIVE condition.

Positive comments in the FLUENCY condition included subjects reporting to be “highly impressed [with the robot's] learning”, and a subject saying that they “had emotional responses that went from tenderness [...] to amusement to respect [...] and trust.” And one went so far as to claim that “[b]y the end of the second sequence, we were good friends and high-fived mentally after the task was done.” Such positive comments were rare in the REACTIVE condition.

Several negative comments, in particular with regards to the robot's contribution as a team member, were found throughout the comments of subjects in the REACTIVE condition. These included “The robot was more of an assistance than an active team member”, and “I felt like I was controlling the robot, rather than it being part of a team,” and “[...] it just felt like a lazy apprentice”. This also reflected on the overall sense of the team's accomplishment, in remarks such as “I'm not sure our team performance ever improved.” In contrast, subjects in the FLUENCY condition remarked on the robot's contribution to the team, and referred to it several times as a teammate: “By the end of the first sequence I realized that he could learn and work as my teammate”, “my interaction with the robot was not that different than with a human teammate,” and I sometimes believed [the robot] a better performer than myself and was impressed at the rate of improvement we had”. One subject in the FLUENCY condition said, however, that “I did not perceive it as human but more as kind of a thing, possibly an animal. I think this might have to do with the fact that I was asked to give it short commands similar to the ones given to animals.”

#### 6.4.2 Self-deprecation in the FLUENCY condition

A surprising effect of the experiment was that in the FLUENCY condition we found a high number of self-deprecating

comments, and comments indicating worry or stress of fallible human performance in relation to the robot's strong performance. Several subjects in that condition remarked on stressful feelings that they weren't performing at an adequate level.

These remarks included "I would essentially forget the pair of colors I had [memorized]—this slowed me down", "The robot is better than me", "The performance could had been better if I didn't make those mistakes", "[I] worried that I might slow my teammate down with any mistakes I might have made", and even "I am obsolete". There were no similar comments in the REACTIVE condition.

While it is beyond the scope of this work to further explore this aspect of our findings, it should be of interest to designers in the human-robot interaction field. The prevalence of this reaction may indicate a need for humans to feel more accomplished than the robot they are interacting with. Maintaining the balance of increased robot responsiveness, and the intimidation that might result is an overlooked aspect of HRI, which these results urge us to consider.

#### 6.4.3 Lexical analysis

We confirmed these anecdotal findings using an independent qualitative coding of the open question responses. All-in-all there were 54 comments, 24 by subjects in the REACTIVE condition, and 30 by subjects in the FLUENCY condition. The comments were presented to a coder in randomized order and without indication of the condition they belonged to. For each comment, and for each category, the coder was asked to answer whether the comment has that property. Specifically, subjects in the FLUENCY condition commented on the robot more positively, and subjects in the REACTIVE condition commented on the robot more negatively. FLUENCY subjects attributed more human characteristics to the robot, although there is little difference in the emotional content of the comments. Subjects in the FLUENCY condition tended both to give more credit to the robot, and attribute more blame to themselves. Those in the REACTIVE condition far exceeded the other subjects in putting blame on the robot. Also, gender attributions, as well as attributions of intelligence occurred only in the FLUENCY condition, while subjects in the REACTIVE conditions tended to comment on the robot as being unintelligent. Finally, we did confirm the tendency to self-deprecating comments as more prevalent in the FLUENCY condition.

## 7 Conclusion and future work

For robots to be accepted as productive members of human-robot teams, they must be able to act fluently with a human

partner in real-world situated teamwork scenarios. These robots must overcome strict turn-taking behavior, which induces delays and inefficiencies, and can cause frustration. Instead they should mesh their behavior dynamically with their human counterparts. This is particularly true for a repetitively practiced joint task, where human teammates have been shown to expect an increasingly coordinated interaction with the robot, even when not prompted to expect this behavior (Hoffman and Breazeal 2007).

In this paper, we introduce a novel cognitive architecture aimed at achieving such fluency in human-robot joint action. Based on neuro-psychological findings in humans, we propose a perceptual symbol system, which uses anticipatory simulation and Hebbian inter-modal reinforcement to decrease reaction time through top-down biasing of perceptual processing along efferent processing pathways.

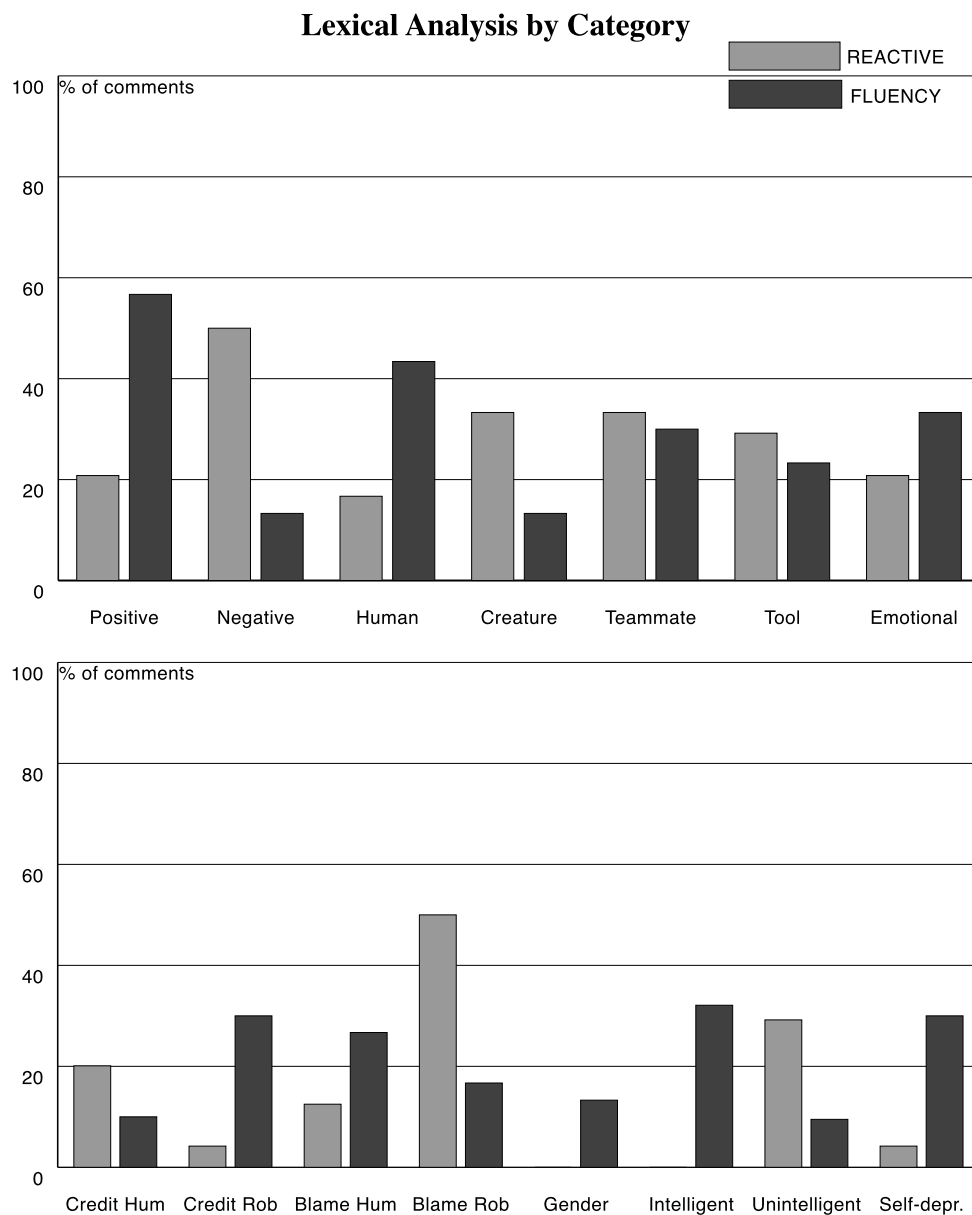
We present a human subject study evaluating the effects of our approach, comparing it with a similar system using only bottom-up processing. We evaluated our system in a repetitive task in which one human and one robot work together using distinct actions to achieve a common goal.

From behavioral analysis, we find significant differences in the task efficiency and fluency between the two conditions. We find the team including the anticipatory robot to be more efficient in all evaluated scales, and its behavior to match our previous established fluency metrics, such as human idle time and robot functional delay. This supports our hypothesis that top-down anticipatory perceptual simulation can aid in fluent human-robot teamwork in which a human and a robot jointly practice a task.

We have also evaluated the relative improvement of the human and the robot, a hitherto under-addressed metric of human-robot joint activities. In our experiments, we find a similar learning curve for the human and the robot team members, possibly contributing to the human subjects' sense of similarity to the robot. We believe that this finding affords a discussion and additional research relating to the psychological effects of human-robot co-learning. For example, we would like to evaluate the bonding effect that result from faster vs. slower adaptation on the robot's part. Do humans generally prefer robots that learn at a similar rate as they do? Does a faster or a slower learning rate frustrate human team members? Does artificially matching the robot's learning rate to that of a human collaborator contribute to the human's sense of likeness or to the bond between human and robot team members? We leave these questions to future work.

From self-report, we find significant differences in a number of metrics, in particular in the perception of the team's fluency, the team's improvement over time, the robot's contribution to the efficiency and fluency, the robot's intelligence, and in the robot's adaptation to the task.

**Fig. 14** Lexical analysis of subjects' open-ended comments. For numerical results, see Hoffman (2007)



In open questions, we find significant differences in the subjects' attitude towards the robot: most notably an increased attribution of human qualities to the robot, such as gender and intelligence, as well as credit for success. Interestingly, we also find a tendency towards self-deprecation in subjects collaborating with the anticipatory version of the robot. This finding, too, affords further research in the area of self-image of humans acting with adaptive robots. The detrimental effects of robot efficiency and adaptation have so far not been sufficiently addressed by the literature. We plan to further evaluate self-image of humans working with robots as it relates to the robot's performance.

A number of further open questions remain.

As the robot becomes more proficient at the task, its primed actions have an increasing effect compared to

sensory-originating activity. While this enables the robot to become increasingly automatic, it could also interfere with the robot's attention to unexpected events, potentially making it too primed, and thus blind to sensory data. In our implementation, we have tuned the system to never completely ignore sensory input, by setting the simulation/activation rate so that any motor activity necessitates some degree of sensory confirmation. Also, in our model, perceptual simulation decays and, if not supported by sensory data, is swiftly overridden by real-world perception. The learning rate, the temporal dynamics (i.e. the onset and decay of perceptual simulation), and the dynamic mixture model of simulated and real perceptual activation is a fertile ground for exploration, and will be further investigated as part of this research.

In future research, we would also like to address the role-dynamics throughout a collaborative session, and measure the attention given to the robot as it changes throughout the session. An interesting question would be whether trust in the robot alters the human partner's measurable behavior, resulting in less supervision over the robot, less confirmatory nonverbal behavior, and less coordination acts, as we would expect from humans that trust each other to do a task correctly. We plan to measure these metrics in follow-up experiments.

In addition, our human subjects' a-priori notions of robots seem to have a conflating effect on their subjective evaluation of the robot's traits and performance. We would like to investigate ways to control for these cultural expectations, given that our goal is long-term engagement with personal robots.

Finally, this work has evaluated our system over a single session, in which the task was unknown to the robot. It is valuable to extend the application of the principles set forth herein over multi-task sessions. In particular, we would like to explore the transfer of adaptation to identical tasks, to changing tasks, and to tasks with new human team members. Also, the application of our framework to known task structures affords further research. Known structure could feed the perceptual simulation system as a third subsystem representing long-term memory, in addition to the short-term memory model described herein. Our framework lends itself to these extensions, as learned behaviors from one task could be used in subsequent tasks. That said, the effect it has on human team members is subject to empirical evaluation.

In conclusion, this work presents steps towards our larger goal of modeling artificial practice in the context of human-robot collaborative tasks, and of building robots that can improve in fluency through repetitive situated practice. We have approached these goals by proposing a novel computational model for joint human-robot team practice, and by evaluating this model in a joint activity study involving untrained human subjects.

## References

- Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Bregler, C. (1997). Learning and recognizing human dynamics in video sequences. In *CVPR '97: proceedings of the 1997 conference on computer vision and pattern recognition* (p. 568). Los Alamitos: IEEE Computer Society.
- Duffy, B. (2000). *The social robot*. PhD thesis, University College Dublin, Ireland.
- Endo, Y. (2005). Anticipatory and improvisational robot via recollection and exploitation of episodic memories. In *Proceedings of the AAAI fall symposium*.
- Fong, T. W., Thorpe, C., & Baur, C. (2001). Collaboration, dialogue, and human-robot interaction. In *Proceedings of the 10th international symposium of robotics research*, Lorne, Victoria, Australia. London: Springer.
- Hamdan, R., Heitz, F., & Thoraval, L. (1999). Gesture localization and recognition using probabilistic visual learning. In *Proceedings of the 1999 conference on computer vision and pattern recognition (CVPR '99)* (pp. 2098–2103), Ft Collins, CO, USA.
- Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory*. New York: Wiley.
- Hoffman, G. (2007). *Ensemble: fluency and embodiment for robots acting with humans*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Hoffman, G., & Breazeal, C. (2004). Collaboration in human-robot teams. In *Proceedings of the AIAA 1st intelligent systems technical conference*. Chicago: AIAA.
- Hoffman, G., & Breazeal, C. (2006). Robotic partners' bodies and minds: an embodied approach to fluid human-robot collaboration. In *Fifth international workshop on cognitive robotics (AAAI'06)*.
- Hoffman, G., & Breazeal, C. (2007). Cost-based anticipatory action-selection for human-robot fluency. *IEEE Transactions on Robotics and Automation*, 23(5), 952–961.
- Hoffman, G., & Breazeal, C. (2008). Achieving fluency through perceptual-symbol practice in human-robot collaboration. In *Proceedings of the ACM/IEEE international conference on human-robot interaction (HRI'08)*. New York: ACM.
- Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the working alliance inventory. *Journal of Counseling Psychology*, 36(2), 223–233.
- Jones, H., & Rock, S. (2002). Dialogue-based human-robot interaction for space construction teams. In *IEEE aerospace conference proceedings* (Vol. 7, pp. 3645–3653).
- Khatib, O., Brock, O., Chang, K., Ruspini, D., Sentis, L., & Viji, S. (2004). Human-centered robotics and interactive haptic simulation. *International Journal of Robotics Research*, 23(2), 167–178.
- Kimura, H., Horiuchi, T., & Ikeuchi, K. (1999). Task-model based human robot cooperation using vision. In *Proceedings of the IEEE international conference on intelligent robots and systems (IROS'99)* (pp. 701–706).
- Marsella, S., & Gratch, J. (2009). EMA: a process model of appraisal dynamics. *Cognitive Systems Research*, 10(1), 70–90.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70–76.
- Simmons, K., & Barsalou, L. W. (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology*, 20, 451–486.
- Spivey, M. J., Richardson, D. C., & Gonzalez-Marquez, M. (2005). On the perceptual-motor and image-schematic infrastructure of language. In Pecher, D., & Zwaan, R. A. (Eds.), *Grounding cognition: the role of perception and action in memory, language, and thinking*. Cambridge: Cambridge University Press.
- Ude, A., Moren, J., & Cheng, G. (2007). Visual attention and distributed processing of visual information for the control of humanoid robots. In Hackel, M. (Ed.), *Humanoid robots: human-like machines* (pp. 423–436). Vienna: I-Tech Education and Publishing.
- Walker, W., Lamere, P., Kwok, P., Raj, B., Singh, R., Gouvea, E., Wolf, P., & Woelfe, J. (2004). *Sphinx-4: a flexible open source framework for speech recognition* (Tech. Rep. TR-2004-139). Sun Microsystems Laboratories.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636.
- Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131, 460–473.
- Woern, H., & Laengle, T. (2000). Cooperation between human beings and robot systems in an industrial environment. In *Proceedings of the mechatronics and robotics* (Vol. 1, pp. 156–165).
- Wren, C., Clarkson, B., & Pentland, A. (2000). Understanding purposeful human motion. In *Proceedings of the Fourth IEEE international conference on automatic face and gesture recognition* (pp. 378–383).



**Guy Hoffman** received his M.Sc. in Computer Science from Tel Aviv University, and his Ph.D. from the Massachusetts Institute of Technology (MIT) Media Laboratory. He is currently a postdoctoral research fellow at the Georgia Institute of Technology Center for Music Technology. His research investigates practice, anticipation, and joint action between humans and robots, with the aim of designing personal robots that display more fluent behavior with their human counterparts.

Other research interests include theater and musical performance robots as well as non-anthropomorphic robot design



**Cynthia Breazeal** received the B.S. degree in electrical and computer engineering from the University of California, Santa Barbara, and the M.S. and Sc.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 1993 and 2000, respectively. She is an Associate Professor of Media Arts and Sciences at MIT. Her interests focus on human-like robots that can interact, cooperate, and learn in natural, social ways with humans.