[PREPRINT]
Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

Robot feedback shapes the tutor's presentation. How a robot's online gaze strategies lead to micro-adaptation of the human's conduct

Karola Pitsch, Anna-Lisa Vollmer & Manuel Mühlig

Abstract

The paper investigates the effects of a humanoid robot's online feedback during a tutoring situation in which a human demonstrates how to make a frog jump across a table. Motivated by micro-analytic studies of adult-child-interaction, we investigated whether tutors react to a robot's gaze strategies while they are presenting an action. And if so, how they would adapt to them. Analysis reveals that tutors adjust typical "motionese" parameters (pauses, speed, and height of motion). We argue that a robot – when using adequate online feedback strategies – has at its disposal an important resource with which it could pro-actively shape the tutor's presentation and help generate the input from which it would benefit most. These results advance our understanding of robotic "Social Learning" in that they suggest a paradigm shift towards considering human and robot as one interactional learning system.

Keywords: Human-Robot-Interaction; Feedback; Adaptation; Multimodality; Gaze; Conversation Analysis; Social Learning; Pro-active Robot Conduct

1. Introduction

If at some point robotic systems (and other autonomous technologies) were to be deployed in everyday life situations, they would need to be equipped with a means for flexible adaptation to new situations and tasks. In this context, researchers strive to develop mechanisms that make it possible for lay users to teach a system new behaviors by way of ordinary language and interaction. Within this "Social Learning" paradigm,

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

tutoring and imitation scenarios play an important role: a human tutor presents and explains a task to a robot, who is then supposed to observe the human, understand the action and, in turn, attempt to reproduce it (Breazeal & Scassellati 2002; Wrede et al. 2008; Cangelosi et al. 2010). As such, beyond sophisticated online learning algorithms, success also depends on the quality and nature of the tutor's presentation. While one line of research focuses on advancing methods for detecting and analyzing the tutor's performance, we suggest the importance to further explore the ways in which the robot could best exploit the interaction with a human tutor.

Most existing human-robot-interaction (HRI) studies on "Social Learning" consider the robot a *passive* observer of the situation. However, in human-computer-interaction (HCI) the relevance of the system's feedback to display its internal status and how it is programmed is well established. Research on human social interaction allows for more fine-grained insights and shows that participants monitor each other and – based on their online-analysis – attempt to closely co-ordinate their actions with those of their co-participant (e.g. Mondada 2006). When adult tutors present and explain a manipulation action to their infant, such as stacking differently sized cups, they adjust the movement of their hands in step with the infant's changing visual focus of attention (Pitsch et al. 2009, submitted). In this way, the tutor's emergent action presentations and resulting hand trajectories are interactively co-produced.

In this paper, we use these observations to motivate an investigation into whether a robot's *online* feedback can *pro-actively* shape the tutor's action presentation, and if so, how the robot can help generate the input from which it would benefit most. We address these questions by looking at tutors' reactions to a robot's gaze strategies while they are presenting an action. In particular, we focus on any adaptations, if any, the tutors make

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

in response to the robot's behavior. We consider the human and robot dyads as an interactional system that uses the human's ability to flexibly adapt to the situation and to his co-participant – a resource largely unexplored in HRI. While learning approaches in robotics tend to investigate the system's learning mechanisms, we turn the question around by asking how we need to design a robot's online feedback so that the human tutor can best make use of his adaptational capabilities, and offer the input most suitable to the robot's learning mechanisms.

In what follows, we introduce the background on adaptation and co-ordination in "Social Learning" (section 2), detail the setup and design of the HRI-study (section 3) and the analytical method (section 4). Section 5, 6 and 7 present fine-grained analyses of a collection of cases with different sets of robot online feedback. We draw further implications for the design of robot behavior in "Social Learning" scenarios (section 8).

2. Adaptation and co-ordination in Social Learning

2.1 Adult-Child-Interaction: Scaffolding and multimodal co-ordination

According to the socio-constructionist approach, 'learning' is a social endeavor rooted in the situated and communicational practices of collaborating co-participants (Wertsch 1985; Fogel 1993). Often an expert/tutor helps the novice/learner to understand new actions (Gergely & Csibra 2005) and attempts to provide support tailored to the learner's specific needs (Zukow-Goldring & Arbib 2007). In doing so, the tutor adjusts his presentation to the learner's displayed abilities and state of understanding ("scaffolding", Bruner 1985; Vygotsky 1978) and e.g. gradually reduces the support as the learner's ability to perform a given task increases (Pea 2004). Research on second language acquisition has shown how a link between the socio-constructionist

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

perspective and interactional approaches, such as Conversation Analysis, provides insights into the communicational procedures by which participants create suitable learning conditions (Mondada & Pekarek-Döhler 2000, Dausendschön-Gay 2003).

The communicational processes in adult-child-interaction are particularly interesting for robotic "Social Learning". Comparable to very young infants, robotic systems also have limited perceptual and cognitive abilities. This leads to the hypothesis that tutors might deploy similar communicational resources when scaffolding their actions for their respective recipient groups (Rohlfing et al. 2006; Zukow-Goldring & Arbib 2007). Parents carefully modify their speech when tutoring their young infants ("motherese") as well as their actions ("motionese") to render specific aspects of the presentation more salient (Fernald & Mazzie 1991; Brand et al. 2007; Rohlfing et al. 2006). Particular "motionese" features have been revealed which indicate that parents make longer pauses between subactions, present the action more slowly ('velocity'/'pace') and with more exaggerated movement ('range') when interacting with their infants as opposed to with other adults (Vollmer et al. 2009).

Taking these observations further, Pitsch et al (2009, submitted) suggest an interactional account of "motionese". Differentiating between the infant's online feedback, i.e. during the tutor's action presentation, and turn-by-turn feedback, i.e. after an utterance/action (Vollmer et al. 2010), they explored the fine-grained interplay between tutor and learner during the action presentation. Based on the participants' mutual monitoring, an interactional loop between the tutor's hand motions and the infant's gaze was revealed. When presenting a manual action, the tutor attempts to guide the infant's visual attention by adjusting the movement of his hand. In turn, the learner's gaze (following/anticipating the action, disorienting) pro-actively shapes the emerging

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

trajectory of the tutor's hand. In particular, cases in which the infant's gaze anticipates the next action are interesting: Tutors treat the infant's anticipating gaze either as a display of lack of attention by responding with a more salient motion, i.e. a particularly high action trajectory, or they treat it as a display of understanding, downgrading their presentation to a flat movement. In this paper, we build on these findings and use them as motivation, for designing and exploring feedback strategies in a robotic learner.

2.2 Robotic "Social Learning"

Robotic learning approaches have a longstanding tradition in developing algorithms for "offline" learning where the human conduct appears as corpus-based training data. More recently, the social dimension has increasingly been taken into account suggesting that an autonomous system could learn from interacting with the human (Breazeal & Scasselati 2002; Steels & Kaplan 2002). In addition to learning algorithms, the robot also needs to organize and manage the interaction with the tutor, i.e. engage in turn-taking, establish joint attention, ground actions and provide feedback (Wrede et al. 2008). However, so far, imitative learning interactions between human and robot have tended to be characterized mainly as one-way communication, where the robot observes the tutor's actions without actively contributing to the social situation. For example, some studies investigate the tutor's conduct by confronting him with a *static image* of a robot to which he should present some action (Herberg et al. 2008). In those cases where a *dialogic* perspective on imitation learning is taken (Alissandrakis et al 2011), the robot is generally programmed to provide a positive/negative statement *after* the tutor has finished his presentation. In such cases, tutors indeed acknowledged the robot's feedback, but they required it to be more informative.

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

Extending the existing approaches of "Social Learning " in Robotics, we explore the idea of interactional co-construction as it is basic to Conversation Analytic research and has been combined with socio-cognitive theories of human learning (Mondada & Pekarek-Döhler 2000, Dausendschön-Gay 2003). In the course of several studies we have begun to investigate how a robot's *online* feedback can *pro-actively shape* a tutor's action presentation. Pitsch et al (2012) presented a first approach in which an autonomous iCub robot provided online feedback during a tutor's action presentation. It observed the tutor's changing gaze direction and pointing gestures while attempting to reciprocate them. When comparing responsive vs. non-responsive robot behavior, it was found that a robot's conduct during the first twenty seconds of an interaction shaped the way in which the tutor presented the action, thus resulting in different tutoring styles (dialogic vs. monologic presentation format). In cases where an initially responsive robot later produced incoherent behavior, tutors were found to be more forgiving. When responsiveness failures occurred, they were normalized by the tutor if the robot's conduct provided elements that could be interpreted as meaningful, and integrated into the sequential structure of the tutor's presentation. Thus, initial evidence exists of tutors' adaption when reacting to a robot's online feedback. A more detailed examination of what exactly tutors react to, how they interpret the robot's conduct, and what strategies might be beneficial, requires further investigation.

3. HRI-Study: Towards feedback strategies for a robotic learner

3.1 Setup and Task: "Please show the robot how the frog jumps"

We conducted an HRI study in which 59 participants (native German speakers, right-handed, aged 20-60 years, with no previous experience with robots) were asked to act as

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

a tutor to a humanoid robot (Vollmer 2011).[1] Each participant was seated on one side of a table vis-à-vis the robot (Fig. 1a) and was asked to consecutively present a set of 8 actions to the robot involving the manipulation of an object. After each presentation, the robot attempted to reproduce the action and the participant was asked to decide whether the robot's reproduction was satisfactory or not. They could repeat their presentation until they were satisfied with the robot's reproduction.

Figure 1: Setup. **(a)** Robot and Participant facing each other; **(b)** Object 'frog'.

In this paper, we focus on fragments from one particular task that consisted of demonstrating how a toy frog (Fig. 1b) jumped across the table. The participants were instructed as follows: "Please show the robot how the frog jumps". The instructions were purposely underspecified so as to allow the tutors to explain the procedure using a combination of verbal and non-verbal input at their own discretion. We chose this task because it involved a series of comparable tutor motions and visible changes in the robot's head orientation.

3.2 Design of the Robot's Feedback

The robot was equipped with a set of feedback strategies motivated by the "interactional loop" between the learner's gaze and the tutor's hand motions in the adult-child tutoring

[1] The study was conducted at the CoR-Lab Bielefeld as part of the project 'Acquiring and Utilizing Correlation Patterns across Multiple Input Modalities for Developmental Learning' funded by the Honda Research Institute Germany and carried out by Anna-Lisa Vollmer and Manuel Mühlig in collaboration with Karola Pitsch, Katharina Rohlfing, Britta Wrede, Jannik Fritsch and Jochen Steil (see Vollmer 2011).

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

scenario presented in Pitsch et al (2009, submitted). Three different versions of online gaze feedback were implemented, which the robot produced during the tutor's ongoing presentation and which provided a pre-structured collection of interactional cases.

(1) **Action-Related Gaze:** The robot's head is oriented to the action. In (1a) the robot's head follows the tutor's hand motion once the object has been picked up at the start position until it is placed at the goal position on the table (**Following**). In (1b) the robot's head initially follows the tutor's hand motion, but after 2 seconds shifts towards the goal position and thus anticipates the end of the tutor's action (**Anticipating**). These differences were chosen to test the hypothesis that the robot's systematic gaze-following would enable the tutor to perform his action presentation without significant disturbances. In contrast, the combination of following-anticipating was expected to yield some sort of confusion.

(2) **Relevant Random Gaze:** The robot directs its head to five different locations (object, start position, goal position, tutor's face and tutor's stationary hand) in random order and with varying (but realistic) durations. These five locations are the most prominent places to which infants in a comparable situation orient (Pitsch et al 2009, submitted) and are thus relevant to the ongoing action.

(3) **Static Gaze:** The robot's head is fixed towards an intermediary position between tutor and table appearing to have both parts 'in view'.

Additionally, the robot attempts to *reproduce the observed action*. The robot either (i) tries to reproduce the trajectory of the observed action (**Imitation**) or (ii) reproduces the goal of the action without respecting the trajectory covered by the tutor's hand, i.e. it transports the object in a straight line to the goal position (**Emulation**). During the experiment, conditions (2) and (3) were combined with both (i) and (ii), while condition (1a) was combined with (i) and (1b) with (ii) (see below Table 1).[2]

The experimental platform used was the Honda Humanoid Research Robot, a 1.20 m sized humanoid robot set up to run autonomously (Mühlig 2009). To enable the robot to detect and follow the tutor's hand movements and the object's position and trajectories, marker-based tracking methods were used. A Polemus marker was attached to the object and the tutor's hands and head were equipped with rigid bodies recorded with the infrared-based Vicon system (Vollmer 2011).

3.3 Data Set

As the recorded data was primarily targeted towards statistical analyses of the tutor's conduct (Vollmer 2011), the order and combination of tasks and feedback conditions was randomized. However, a qualitative explorative analysis of the data requires a set of structurally comparable cases. Therefore, we focus on one particular task ("please show the robot how the frog jumps") and only when it occurs as the first task in a series of 8 to prevent interference effects from the subsequent tasks. This leaves us with a data set of 9 participants where the tutors presented the action 'frog jumping' as the first item.

Online Feedback (Gaze) / Action Reproduction	**Action-related**		**Relevant Random**	**Static**
	Following	**Anticipating**		
Imitation	**VP18**: 7		VP20: 2 (no gaze)	--
Emulation		**VP02**: 3	**VP19**: 3 VP21: 4	**VP27**: 3

[2] In this analysis, we do not take into consideration the robot's action production. Nevertheless, we should note that for condition (1) a connection is suggested, on the one hand, between the robot's gaze-following of the object and the imitative reproduction of the trajectory (1a-i), and on the other hand, between the robot's anticipation of the tutor's action goal and the robot's failure to reproduce details of the action (1b-ii). This connection has been created for the purpose of the HRI study, but was not part of the observations on adult-child-tutoring presented in Pitsch et al. (2009, submitted).

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

			VP43: 3 **VP51**: 6 VP54: 4 (no gaze)	

Figure 2: Data set 'frog'. For the different conditions, the participant codes (VP) are given together with the number of the tutors' repeated action presentations.

For VP20 and VP54, the participants did not look at the robot when presenting the action. As such, no analytical claims can be made about their reactions to the robot's feedback. Thus, the data set analyzed in this paper contains, with one exception (VP18), only cases in which the robot reproduced the action by transporting the object directly in a straight line to the goal position (Emulation).

4. Method of Analysis

The data analysis is based on Ethnomethodological Conversation Analysis (EM/CA, see Goodwin 2000) to provide insights into the sequential structure of the interaction. This method enables us to investigate the interrelationship between robot's and tutor's actions, and how they respond to each other on the level of sequential structures. Further, it aims to reconstruct the participants' view ("member's perspective"). We explored the user's perception and understanding of the robot's actions, and to what extent they constituted a meaningful, relevant action for the participant.

EM/CA is a qualitative approach consisting of manual analysis, i.e. repeated inspection of video-data and transcription of interactions to uncover the timing and relationship of actions. Its goal is to discover the structural organization, in particular how one action makes a subsequent action contingently relevant. In this way we can account for structurally expected, albeit missing, actions during an interaction. EM/CA is based on a set of assumptions about human communication: task orientation, interactivity and co-

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

construction, mutual monitoring & online analysis, sequentiality, and multimodality (see e.g. Pitsch et al. submitted). This framework invites us to consider 'tutoring' as a collaborative achievement between the tutor and learner (compare 'co-development' in Fogel 1993), and to reconstruct the procedures and methods they deploy jointly to do so.

For exchanges between a human and a robot, notions such as 'interaction' and 'co-production' seem problematic. On the one hand, the actions of a human and a machine are based on different structural expectations. Human interactional conduct is situated, i.e. based on a stepwise process of local sense-making practices which allow the human to flexibly react to the emerging contingencies of an interaction, whereas technical systems follow a pre-specified plan (Suchmann 1987). On the other hand, humans are oriented to the structures of ordinary conversation when talking to a machine ("persistence of communication", Hutchby 2001) and they appear to interpret the machine's actions as being those of an acting co-participant (Latour 1988). An in depth discussion of these issues is beyond the scope of this paper. We focus instead on the surface characteristics of the robot's behavior, and how the tutors interpret it as meaningful sequential actions.

In addition to the video-based manual analysis, we captured the trajectories covered by the object with a semi-automatic 2D motion tracker (Vollmer et al., 2009). The tracker generates a time-stamped list of x and y coordinates defining their position in the video frame. In this way, interactional research can benefit from computational methods and begin to overcome the challenge of capturing ephemeral visual phenomena, such as

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

gestures, actions or body movements. These technically generated annotations were combined with manual transcriptions/annotations using the corpus tool 'Elan'.[3]

5. Action-related Gaze: Tutor's adjustment of pauses, speed and height of the hand motion

The data set contains two interactions where the robot used an action-related gaze strategy. In one case it used the 'Following' paradigm (VP18) and in the other, 'Anticipating' behavior (VP02). We investigate how the tutors react to these forms of online feedback, and to what extent they might be able to shape the tutor's conduct.

5.1 The robot's gaze follows the tutor's action: Adjustment of pauses

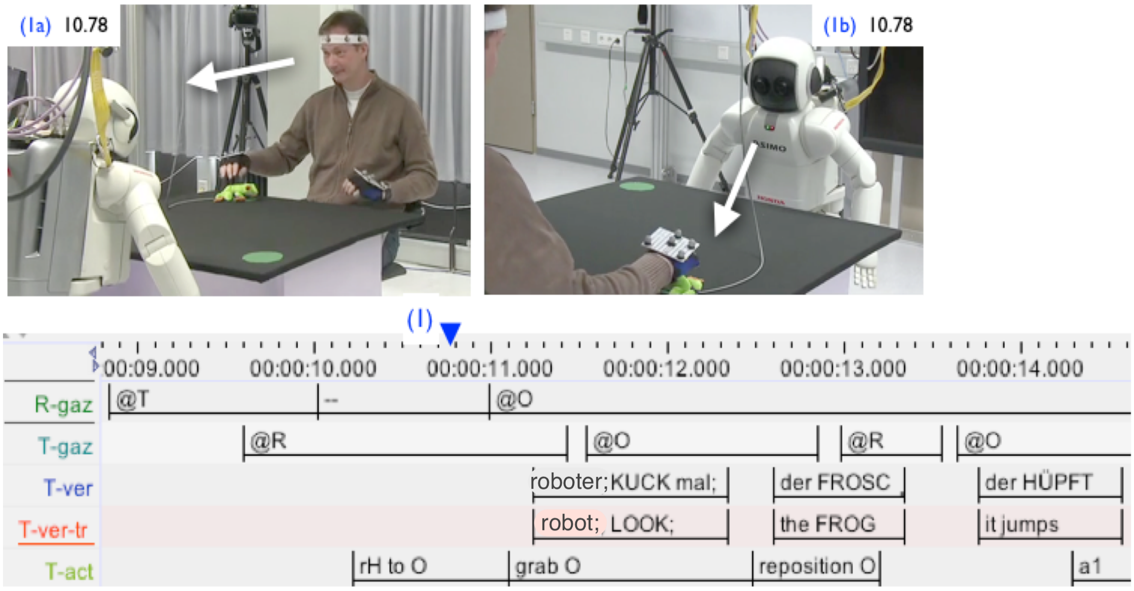
We begin the analysis with a fragment in which the robot was programmed to move its head such that its 'gaze' appeared to follow the tutor's manipulation of the object (condition 1a). Analysis will reveal that the tutor adjusts the duration of his motion pauses in response to the robot's behavior.

For the first fragment, we enter the interaction when the experimenter had just placed the toy frog in the start position on the table. The tutor immediately looks at the frog, then reorients to the robot and reaches forward to take the frog (#10.78). At that moment, the robot also turns its head to the object.

Fragment 1: VP18 – 1st presentation

[3] We gratefully thank Lars Schillingmann for his valuable technical support in combining motion trajectories and video data, and Raphaela Gehle and Lukas Rix for helping with the annotations.

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

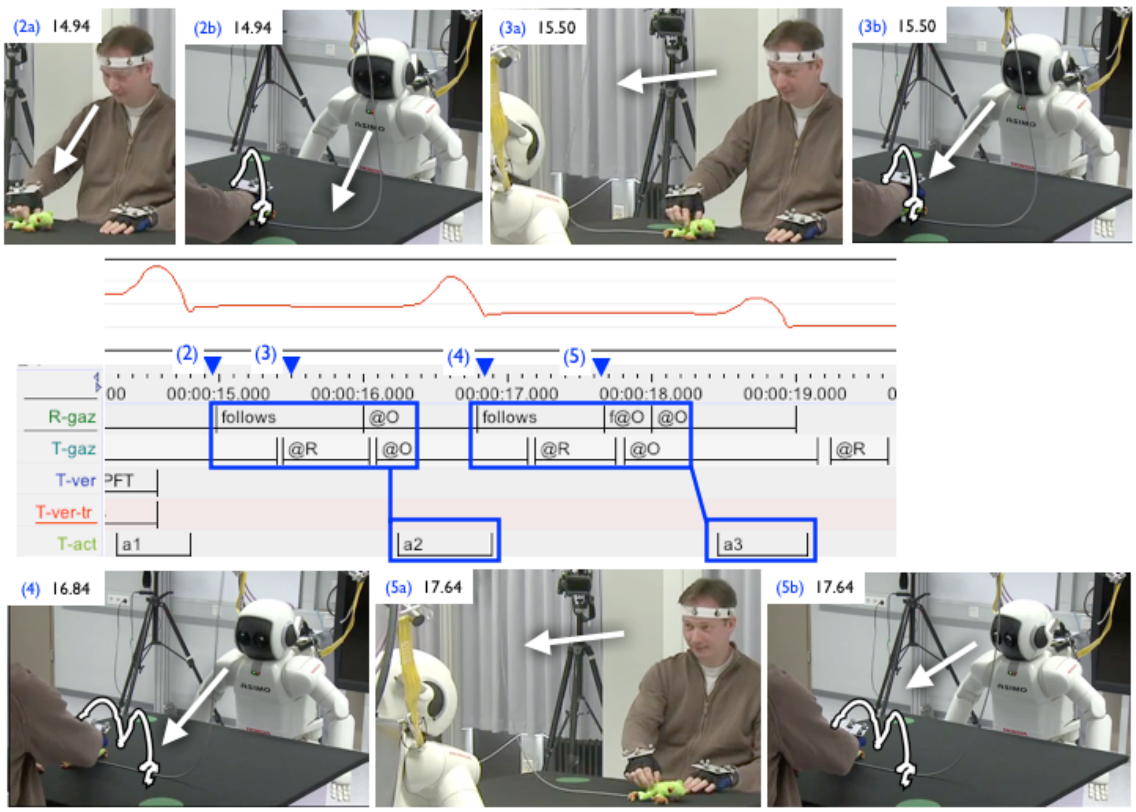


The tutor thus experiences a system that seems to react to changes in the environment and to pro-actively engage in the upcoming activity. This is illustrated by the tutor's subsequent instruction "robot; HAVE a look" (11.20). He treats the robot's changing head orientation as an indicator of an assumed visual observation capacity, and as the system's 'gaze'.

The tutor then demonstrates how the frog jumps. He verbalizes "the FROG, (.) it JUMPS" (12.80 – 14.60) as he takes the frog, lifts and transports it in an arc-like movement a few centimeters across the table (#14.94). This action, transporting the frog, requires that the tutor organize his focus of attention between the object involved (i.e. the frog) and the recipient/co-participant (i.e. the robot). In the present case, he organizes this dual orientation such that he looks at the robot *before* the action (13.00 T-gaz: @R), at the frog *during* the jump (13.60 @O) and again toward the robot *after* the action (15.40 @R). When he observes the robot's behavior after having placed the frog back on the table, he sees that the robot's gaze follows the object's new position (#15.50). It is only once the robot's gaze has reached the frog – i.e. about 1.5 seconds

[PREPRINT]
Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

after the end of the jump action (14.80 > 16.25) – that he continues his presentation. In this way he adjusts the interval (pause) between the first and the second sub-action in step with the robot's behavior.



For the second jump action, we find a similar pattern. The tutor orients toward the frog *during* the jump action (16.10 – 17.20), checks *afterwards* on the robot's conduct (17.20 – 17.75) and sees that its gaze follows the object (#17.64). Again, he only continues with the next jump action once the robot's gaze has caught up (18.50). He continues this pattern for the next presentation of jump actions. Thus, the tutor coordinates his own actions with those of the robot and attempts to establish a coordinated and sequential collaborative action structure. In this way, the robot's gaze co-constructs and shapes the duration of the tutor's motion pauses. He actively influences a typical "motionese" feature.

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

5.2 The robot's gaze anticipates the tutor's action: Pro-actively shaping the emergent action trajectory

In the first fragment, the tutor VP18 did not monitor the robot's conduct *during* his action presentation, only *before* and *after*. This produces a specific condition for the tutor's ability to adapt his actions: he can adjust the moment when he begins a subsequent action in step with the robot's conduct, but not the jump action itself. In the following fragment, the tutor VP02 organized the dual orientation between object and co-participant differently. During the first presentation, he concentrates on the jump motion and ignores the robot. For the 2nd and 3rd presentations, however, he looks at the robot while moving the frog and is thus able to monitor the robot's behavior. In this way, he is not only able to coordinate his conduct in time for the next action; but he also establishes the precondition of monitoring the robot that would allow him to micro-coordinate and potentially adjust his action presentation to the robot's behavior while it is emerging. The tutor's three consecutive presentations provide the opportunity to compare different versions of the same action.

5.2.1 First action presentation: Non-recipient oriented

When the tutor presents the frog jump to the robot for the first time, he takes the frog at the start position, similarly to the tutor in fragment 1, briefly glances at the robot and sees that it has just directed its head to the object (09.80 T-gaz: @R, R-gaz: @Start). He then looks back to the frog and makes it jump in two arcs across the table (#17.94). Thus, during this first action presentation, the tutor is, similarly to VP18, aware of the

[PREPRINT]
Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

robot's initial orientation to changes in the environment, but, in contrast, he presents the action without orienting to the recipient.

Fragment 2: VP02 – 1st presentation

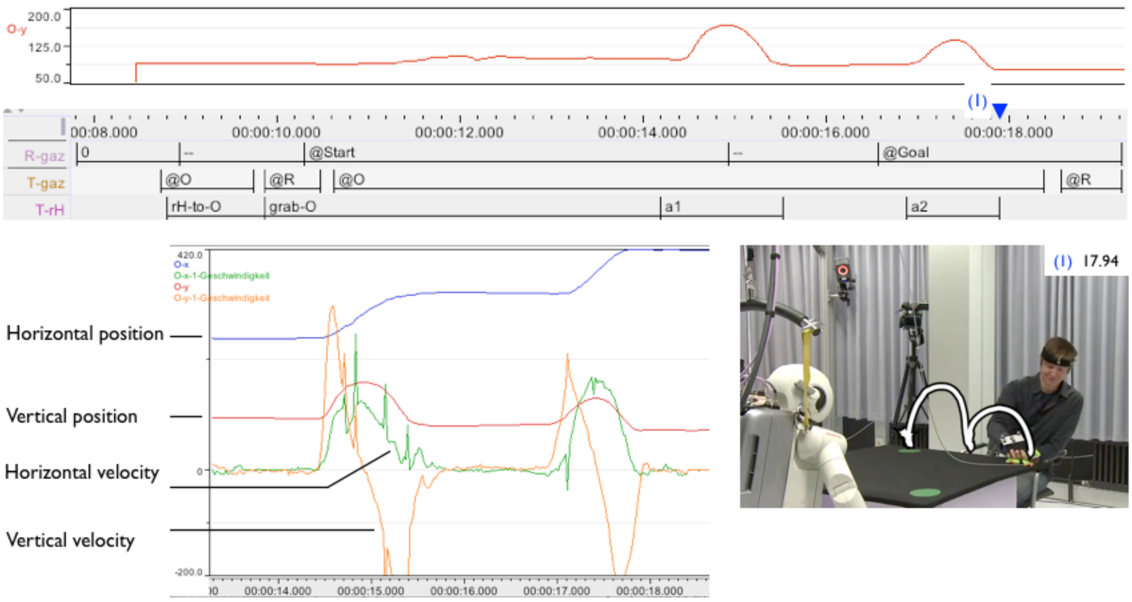


Figure 3: Tutor's 1st (unbiased) action presentation

Using the semi-automatic tracking of the moving object (Fig. 3), we find an even-shaped regular curve[4]: The first line from the top represents the frog's horizontal position as a function of time (x-coordinate of the tracked motion) starting at the tutor's right side and moving across the table to his left side. The second line indicates the frog's vertical position as a function of time (y-coordinate). It exhibits two even-shaped arcs, in which the object's upward and downward movement are symmetrical. The third and fourth lines show the object velocity, i.e. the rate at which the object changes its

[4] If we wanted to undertake mathematical and statistical analysis of the motion data instead of the principled argument here, smoothing and normalizing procedures would need to be applied so that occasional outliers, as is typical for authentic interactional data, would be eliminated. Also, we would rather not track the object's motion on the video data, but use the more sophisticated tracking data recorded in the situation. – We gratefully thank Thomas Hermann for valuable discussions about the physical properties of motion trajectories.

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

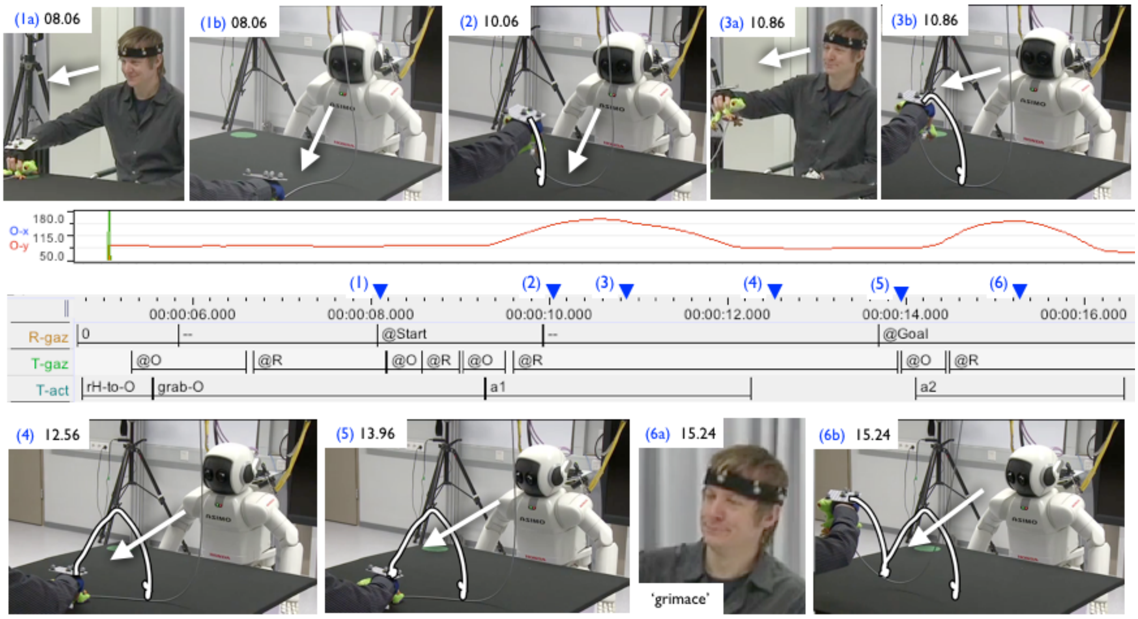
position (in mathematical terms the 1st derivative of the position vector) along the x-axis (third line) respective the y-axis (fourth line). From a physical perspective the trajectory of a frog jumping on an arc is a parabolic curve characterized by a constant velocity along the horizontal axis and a linearly decreasing velocity (from +v to –v) with its typical maximum in the vertical axis. Such trajectories re-occur in the corpus for non-recipient-oriented action presentations and thus can be considered a basic version of the movement.

5.2.2 Second action presentation: Adjustment of motion speed

Given that the robot reproduces the tutor's action by lifting the frog about 10 cm from the table and transporting it in a straight line to the goal position where it is dropped (i.e. without reproducing the jmp motion (emulation, section 3.2)), the tutor decides to present the frog jump again. At the beginning of this presentation, he again gazes at the robot when taking the frog, and notices that the robot's orientation also shifts toward the frog (#08.06). Once the robot's head rests on the object, the tutor exhibits a new gaze strategy. He briefly glances toward the frog (08.20 @O), then back to the robot (08.60 @R), then back to the frog (09.00 @O) and then again to the robot (09.60 @R) as he begins to pick up the frog. While performing the jump action, he therefore monitors the robot's behavior (#10.86) satisfying a basic pre-condition for interactional micro-coordination.

Fragment 3: VP02 – 2nd presentation

[PREPRINT]
Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main



When the tutor has lifted the frog about 10 cm up in the air (#10.06), the robot begins to lift its head to follow the tutor's action presentation within nearly a second of delay (#10.86). The tutor initially observes the robot's action, then begins to slightly adjust his action presentation to the robot's behavior. Although this is difficult to see in video frame captures, despite being very visible in the video itself, the motion tracking data enables these micro adjustments to be examined.

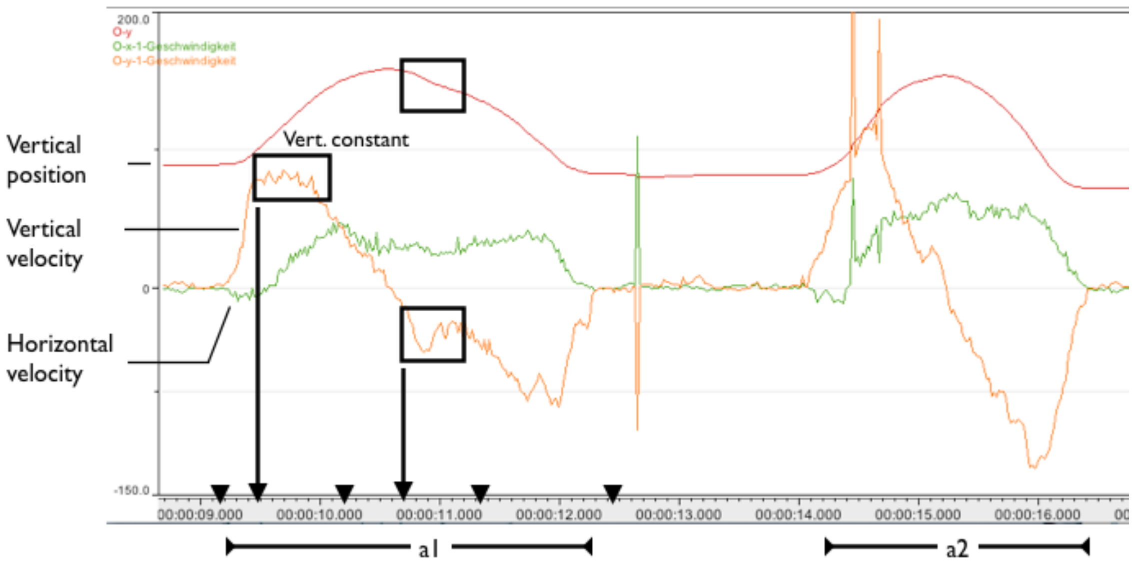


Figure 4: Tutor's 2nd presentation. Deviations from the basic curve are highlighted.

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

At 09.20 the tutor begins to pick up the frog and move it upwards (T-act: a1), which translates as a speed increase in the frog's *vertical* movement (09.20 – 09.50). Being a few centimeters up in the air, at 09.50 the object reaches a constant (i.e. unaccelerated) vertical velocity. In the graph, this produces a linear instead of the basic arc-shaped curve. In the video, this appears as if the object was moved upward in a more straight way in comparison to the curve in the frog's original jump motion. At about 10.10 – i.e. just after the robot has begun to lift its head – the tutor again accelerates the object's vertical velocity and thus returns to the original arc-shaped trajectory. This is shown in the graph as a decrease in the object's vertical speed (note that the object's motion still continues upwards, but due to its maximum point (see Fig. 3) the graph goes downward). Similarly, at about 10.20, the object's *horizontal* movement fades into a constant (i.e. unaccelerated) velocity and thus also resumes the original parabolic curve of the frog jump.

After the arc-shaped curve has reached its peak, the tutor starts to move the object downward from about 10.50 onwards, so that it will eventually meet the robot's rising gaze. When the robot's gaze is about encountering the object (#10.86), the tutor again adjusts around 10.70 to 11.10 the object's vertical velocity. In the object's vertical position (the topmost line) this translates as an 'indentation' in the graph and appears in the video as if the tutor's hand moves down slower attempting to engage the robot's focus of attention to follow the object. Afterwards, starting around 11.30, the tutor's hand movement again resumes the original parabolic curve of the frog jump.

Around 12.40, at the end of this first jumping action (in the transcript: #12.56), the robot's head turns towards the goal position (#13.96) and remains fixed on this location irrespectively of the tutor's further actions. When the tutor continues his presentation

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

with a second jump motion (a2), during which the robot does not show any reactive behavior, the tutor produces a puzzled expression on his face (#15.24). His hand motion, however, continues in a normal, rather unbiased fashion (Fig. 4, 14.00 – 16.50). Thus, it appears that the robot's shifting gaze influences the tutor's action presentation. Not only did the tutor attempt to establish a sequential action structure at the beginning of the next action, but importantly, he also tried to micro-coordinate his hand motions with the robot's gaze behavior. In particular, the adaptation of the tutor's motion speed, i.e. slowing down in relation to the recipient's gaze following, a parameter typical of "motionese" behavior, is thus co-produced by robot and tutor.

5.2.3 Third action presentation: Adjustment of motion speed and height

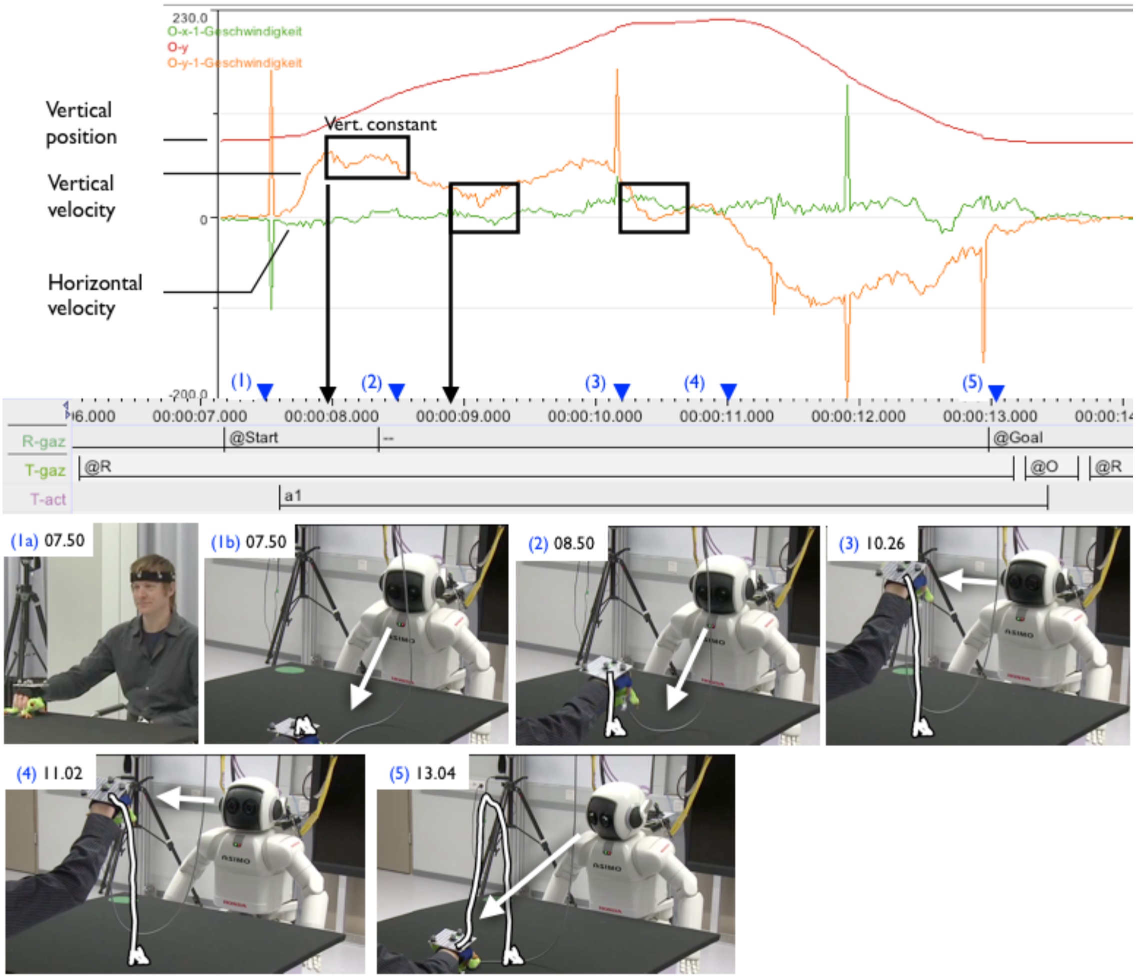
After a second action reproduction by the robot (again reproducing the goal, but not the route) the tutor decided to present the action a third time. Similarly to the previous presentations, he gazes at the robot when picking up the frog, and notices that the robot has shifting orientation toward it. Once the robot's assumed gaze has arrived at the object, the tutor begins to move the frog upwards while continuously monitoring the robot. Within a delay of about 1 second, the robot also begins to raise its head (#07.50 → #08.50). While the tutor's lifting of the object initially translates into an increase of the object's *vertical* velocity (in the graph: the rising line of the vertical velocity, 07.50 – 07.90), from about 08.00 onwards, i.e. with the robot's focus of attention still oriented to the start position in the table, it fades into a non-accelerated motion represented in the graph by the flat line (until 08.50) instead of the expected basic arc-shaped curve. At the same time, the object's *horizontal* verlocity ranges around zero. In the video, these two components take together appear as if the object was moved upward in a rather straight manner (see #08.50).

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

From 08.50 onward, the robot's head follows slowly and the tutor continues to lift the frog, watching as the robot's gaze follows and attempting to micro-coordinate with it. From 08.90 to 09.20, his hand motion nearly comes to a halt waiting for the robot's gaze to catch up. Then, pursuing the coordinated upward movement, at 10.26, the robot's head does not continue its upward motion any further. At this moment the tutor's hand again comes to a near halt (10.30 to 11.00): The vertical verlocity ranges around zero while the horizontal velocity shows a constant, but very low velocity. In the object's vertical position (the topmost line) this translates as a flat line during this time period which is also visualized in the video frame capture #11.02. It is only when the robot's head turns downwards to the goal position, as pre-programmed, that the tutor's hand immediately follows with a downward movement (11.30).

Fragment 4: VP02 – 3rd presentation

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

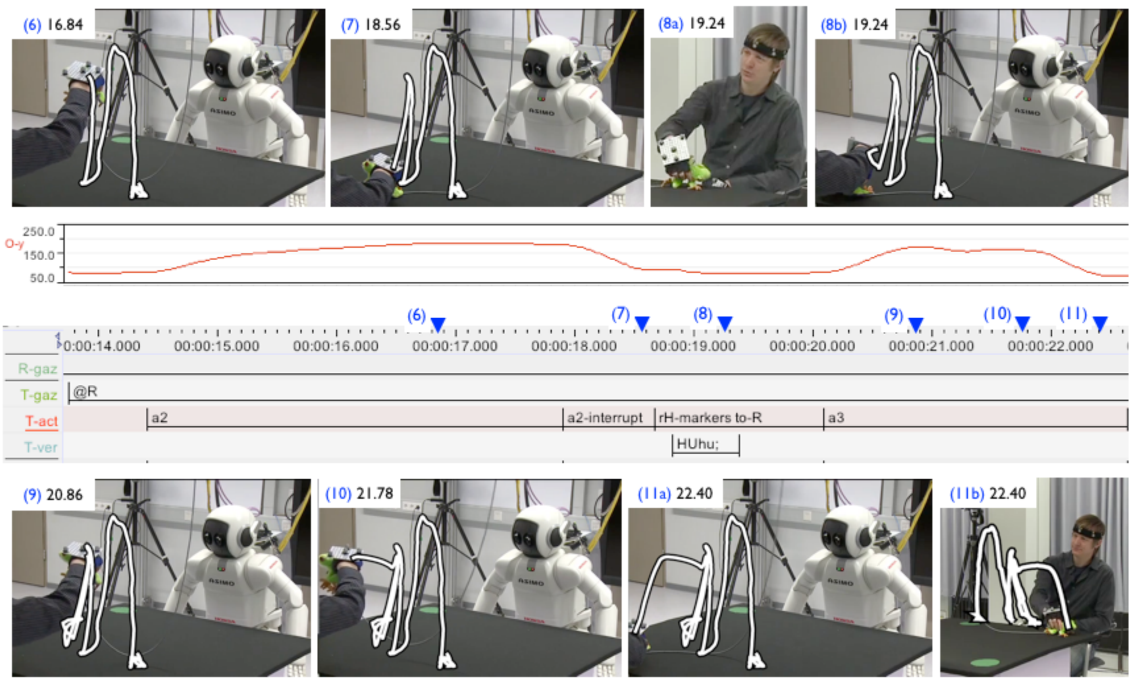


In comparison to the tutor's second action presentation analyzed in section 5.2.2, this sequence reveals even more explicitly the tutor attempts to micro-coordinate his actions with those of the robot. Firstly, the adaptations of the motion speed (slower) are more prominent in that the object's movement not only slows down but also twice nearly comes to a halt waiting for the robot's gaze to follow. The entire action presentation takes significantly more time. While the presentation of the frog jumping took about 1 second in the first presentation, it increased to 3 seconds in the second and to 5.5 seconds in the third action presentation. Secondly, the tutor adjusts the height (higher) of the movement. He raises his hand for as long as the robot's gaze follows to a position of about 200 pixels in the video-frame. His previous motions do not exceed the 160 pixel mark, making the height increase about 20% greater. Thus, this fragment appears

[PREPRINT]
Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

to instantiate also another "motionese" feature, namely, range of the motion.

After briefly positioning the frog on the table, the tutor continues with a second jump. He brings his hand upward. However, the robot's head does not follow but again remains fixed on the goal position (#16.84). The tutor initially slows down his hand motion and stops halfway in the air (#16.58). Then he moves the frog further upward, stops again, interrupts the presentation and places the frog straight down on the table (#18.56). He then attempts to attract the robot's attention by (i) rotating his hand so as to allow the robot an unhindered view of the marker plate on his hand (#19.24), (ii) verbally calling for its attention (18.80 T-ver: "HUhu"), and (iii) moving/waving the frog in the air (#20.86, #21.78). The tutor comments on the robot's failure to focus on the object by saying "if you don't want to look anymore, the frog will only jump once", in this way voicing his interpretation and hypothesis of the robot's function. The apparent lack of visual attention to the second action presentation seems for him to be linked to the robot's failure to reproduce two jumps (as opposed to the single movement which it did reproduce).



Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

In sum, we find evidence that the tutor – when monitoring the robot's conduct during his action presentation – reacts to the robot's online-gaze feedback. He adjusts not only the pauses between different actions, but also the speed and height of the action trajectory to the robot's changing visual focus of attention.

6. Relevant Random Gaze: Integrating the robot's conduct in a relevant sequential structure

The data set contains six cases in which the robot exhibited a random gaze behavior when the tutor was presenting the action. The robot directed its head to five different action relevant locations (object, start position, goal position, tutor's face and tutor's stationary hand) in random order and with varying durations. In two of these instances, the tutor did not visually orient to the robot (and consequently did not show any adaptive conduct). As such, four analytically interesting participants remain (VP19, 21, 43, 51) all of which show a range of similarities. In what follows, we describe two cases in closer detail, and point briefly to parallels found in the other instances. The analysis will reveal that the tutors repeatedly interpret the robot's *random* gaze behavior as being *systematic*. They often explicitly sequence their actions so as to integrate the robot's random behavior into a (for them) meaningful action structure, in this way 'normalizing' the robot's actions. For this interpretation of the robot's conduct, the design of random gaze directed toward *relevant* locations (as opposed to entirely random ones) and with realistic timing, appears to play a substantial role. Importantly, these cases invite us to take human adaptability and sense-making practices seriously as a crucial and highly valuable resource when designing robot behavior for HRI.

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

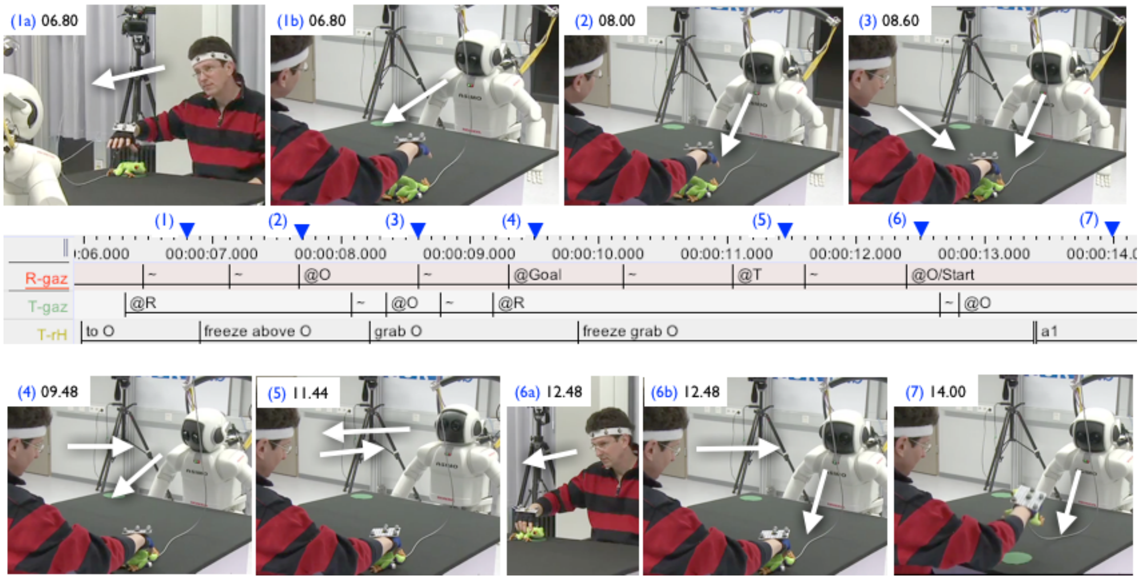
6.1 Normalizing the robot's conduct into meaningful action structures

We investigate the beginning of tutor VP43's third frog jump presentation as an example of the tutors' 'normalization' of the robot's random gaze behavior. During the first and second presentations, the tutor only looked at the object when transporting it. The third action presentation constitutes the first time he closely orients to the recipient. We focus on the beginning of the interaction, just after the frog has been placed on the table. While in the previously examined action-related gaze condition (section 5), the robot was pre-programmed to initially direct its head to the start position on the table once the object had been placed there, the situation differs for the random gaze condition. With the robot's head turning randomly to the five pre-specified locations, the participants in numerous examples can be seen to *actively* attempt to establish co-orientation with the robot toward the object before they begin their presentation.

In fragment 5, after the frog is placed on the table, the tutor brings his hand towards the object while observing the robot. At that point, the robot's head is directed to the opposite side of table. The tutor reacts by stopping his action and freezing his hand in mid-air above the frog (#06.80). Once the robot turns its head towards the frog (#08.00), the tutor continues his action and picks up the frog (#09.48). He then continues to observe the robot, watching as it again reorients, first looking to the goal position (#09.45), and then to the tutor's face (#11.44). The tutor again freezes his action, this time taking the frog, and again, only restarting the presentation once the robot has re-oriented to the frog. (12.30 R-gaz: @O/Start, #12.48, #14.00).

Fragment 5: VP43 – 3rd presentation

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main



These repeated interruptions of an action at the moment when the robot's gaze drifts away, followed by resuming the action once its gaze returns show the systematic character of the tutor's conduct. Such clues suggest that the tutor explicitly (although probably unconsciously) orients to the robot's shifting head movements and gaze directions. Also, participants actively treat the robot's orientation to the frog as a pre-condition to begin the action presentation. Comparable action delays can be found at the beginning of a range of other action presentation (e.g. VP19_03, VP43_2). In particular, these instances show the tutors' approach to understanding the robot's behavior as meaningful actions within an interactional framework. In the random gaze condition, tutors not only acknowledged the robot's initial head orientation to the object (as in section 5), but use it to actively organize their own actions. The tutors explicitly sequence their actions so as to integrate the robot's random behavior into a (for them) meaningful emerging action structure. The same can be found at other moments, when the tutors respond to the robot's behavior as if it constituted further meaningful actions: (i) the robot directing its head to the tutor's face is sometimes responded to with a smile, suggesting a more 'social' quality. Other head movements are understandable as

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

(ii) searching for an object on the table, (iii) as anticipating/introducing the next action or (iv) following the ongoing action. These observations support and expand the analyses presented in Pitsch et al. (2012), where the tutors attempted to normalize a different robot's behavior if they had initially experienced the system as being responsive.

6.2 Further interactional conditions for (non-)adaption

Based on the observation that tutors attempt to make sense of the robot's random actions, to establish meaningful sequences, and coordinate their actions with the robot's behavior, we investigate further instances of the tutors' action presentations in the random condition. This condition complements the designed action-related robot behavior (section 5) in that it provides the opportunity to enlarge the collection of cases and thus produces a range of new interactional situations, allowing further study of the conditions under which tutors may adapt to a robot's behavior.

6.2.1 Action-final small robot head motions do not invite the tutor's adaptation

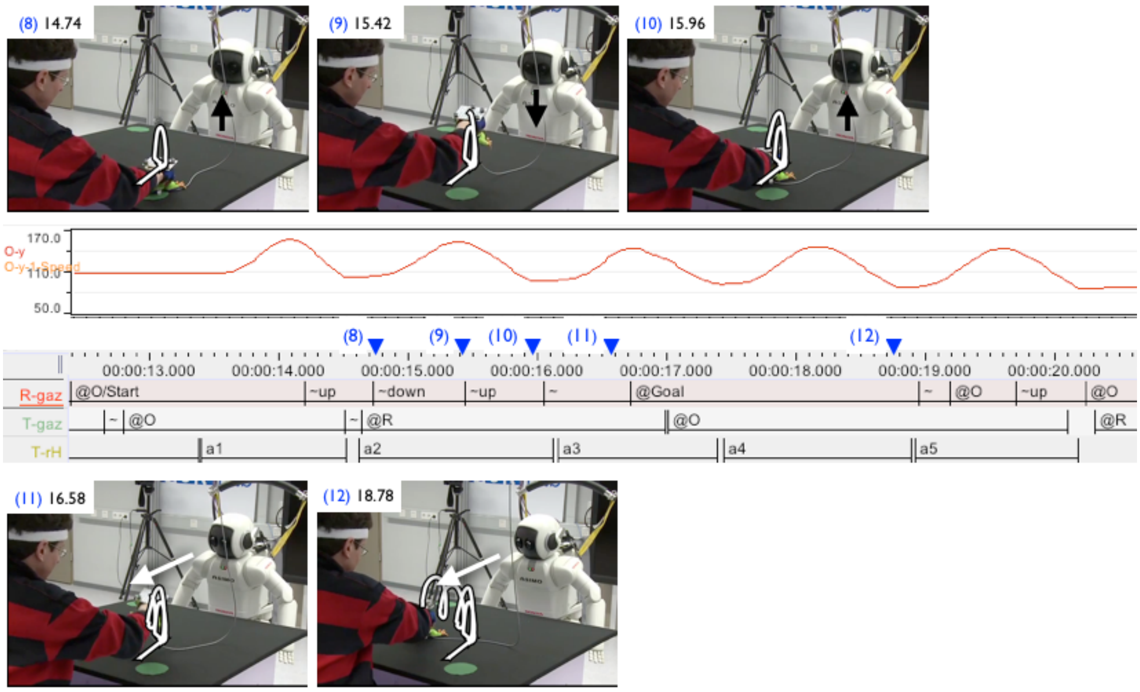
We continue the analysis of fragment 5 (section 6.1), where the tutor postponed the start of his action presentation as an adjustment to the robot's shifting head orientation. We thus have an instance of a highly attentive tutor who actively attempts to co-ordinate his actions with that of the recipient, providing a good basis to further investigate the conditions under which he might adapt to the robot's behavior.

In fragment 6, when the tutor VP43 performs the first jump action, the robot's gaze is oriented to the start position. Just before the frog again lands on the table, the robot's head begins a small up-down movement (#14.74, #15.42). Towards the end of the

[PREPRINT]
Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

tutor's second jump movement, again the robot's head moves up slightly (#15.96). The tutor, however, does not adjust his action trajectory. During the third jump, the robot's gaze shifts directly to the goal, thus anticipating the tutor's actions (#16.58). Again, the tutor continues the flow of his jump movements without visible modifications (#18.78). Similarly to the subsequent jumps, the tutor does not coordinate his actions with the robot's head motions.

Fragment 6: VP43 – 3rd presentation



This example suggests that small up-down-movements of the robot's head do *not* elicit an action modification in the tutor's presentation when they are neither directly related to the trajectory covered by the tutor's hand nor temporally relevant. Also, the timing of the robot's head movement with regard to the stage in the tutor's action presentation seems relevant. Here, the robot's head movements at the *end* of the tutor's hand movement appear less effective for eliciting adaptations than those at the beginning.

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

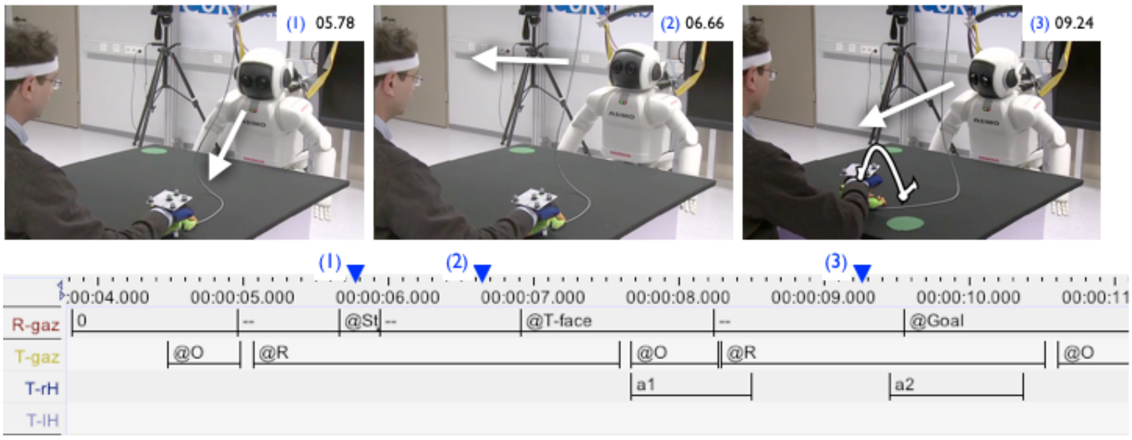
6.2.2 The robot's gaze pre-configures: Co-constructing high action trajectories

In contrast to the previous fragment, in other instances the robot's behavior provides a basis to pre-configure and shape the tutor's presentation by pro-actively lifting its head. In the example shown in fragment 7, the robot's head movement is not entirely systematic, although on occasion it tends to look up to about the same level as the tutor's face. This is treated by the tutor VP51 as a suggestion for a relevant 'interactional space' to perform his presentation. In fact, he reacts by adjusting the height of his action and by coordinating the intervals (pauses) between sub-actions with the robot's head movements.

When we enter the interaction, the tutor picks up the object from the table while simultaneously observing the robot. He sees that the robot directs its gaze to the object (#05.78). Then the robot looks up to the tutor's face (#06.66). The tutor performs the first jump, at the end of which the robot's head re-orients down to the frog (#09.24). The tutor releases the frog and only picks it up again once the robot's gaze has returned to it (another instance of an adjustment of pause duration between sub-actions).

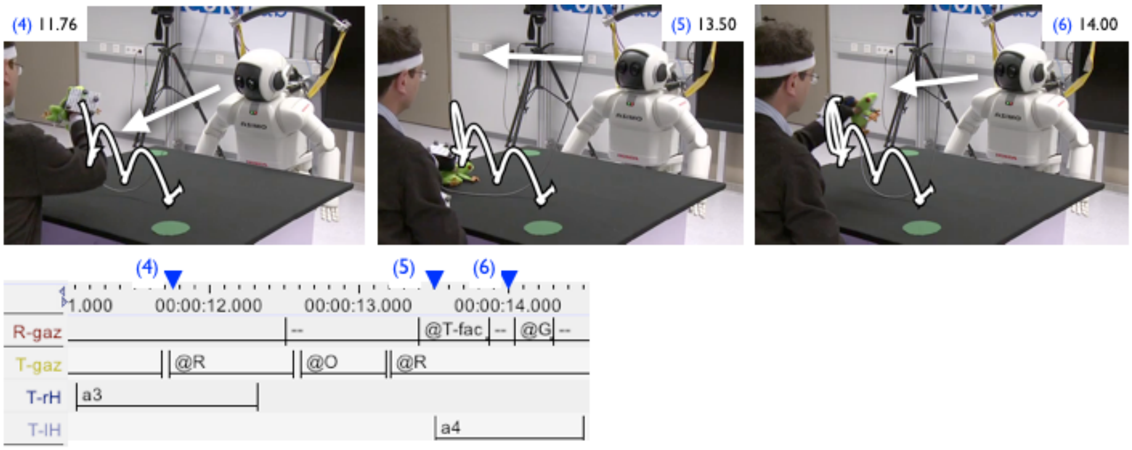
Fragment 7: VP51 – 2nd presentation

[PREPRINT]
Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

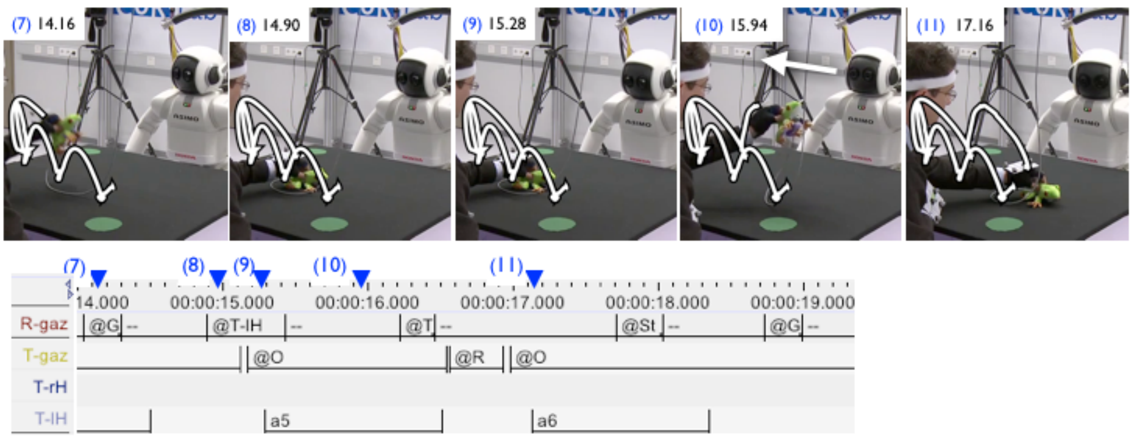


During the second jump, the robot's head remains fixed on the previous location, which is closely monitored by the tutor. During the third jump, with the robot's gaze still fixed on the same location, the tutor performs a *high* jump motion with the frog (#11.76). While the robot does not immediately react, it does so after the tutor has positioned the frog on the table and is about to pick up the frog with the other (left) hand. It again lifts its head to the level of the tutor's face (#13.50) – similarly to its initial behavior (see #06.66). This succession of actions suggests that the robot might be reacting to the rising hand motion. The tutor repeats this new form for the next, fourth jump, again lifting the frog up high. The robot's gaze shifts down until it meets the tutor's hand in mid-air (#14.00). In this way, the tutor's high action trajectory appears (to the tutor) as a co-production between him and the robot across several interactional steps. The robot initially suggests the relevance of gazing high (#06.66), the tutor adopts this as a strategy to activate the robot after it did not react to the action presentation (#11.76). Since the robot appears to be responsive to the high action trajectory (#13.50), the tutor continues with this particular performance (#14.00).

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main



Also, during the course of the action, the tutor again adjusts the pause duration to the robot’s gaze behavior between two sub-actions. When the tutor continues the downward motion of the fourth jump, the robot continues to orient downward, although it first looks left, then right before finally landing again on the frog (#15.28). The tutor waits with his hand (holding the frog) on the table until the robot’s gaze has come to rest on the frog. Similarly, during the fifth jump, the tutor again performs a high action trajectory, with the robot’s head following (#15.94), and continuing even farther up.



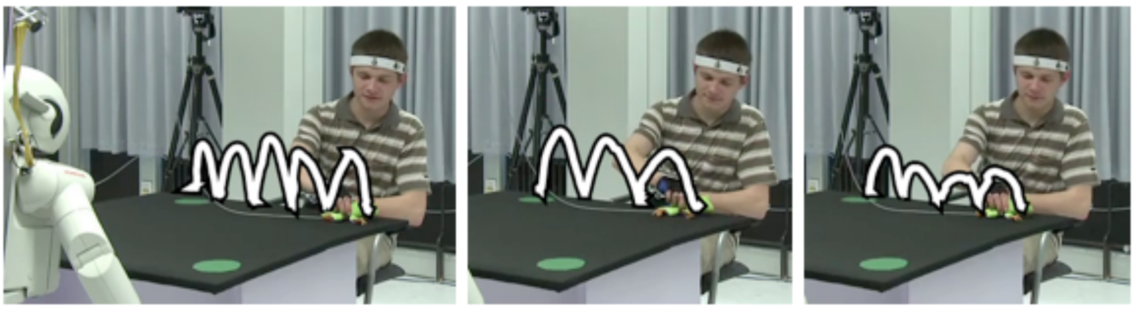
In this way, robot and tutor appear to establish an interactional routine. It seems that a robot could use ‘lifting its head high up’ as a strategy to invite the tutor to also perform the action presentation with a *high* action trajectory and thus pre-configure the ‘interactional space’ for the tutor’s actions.

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

7. Static Gaze: No adaptation by the tutor

In the condition 'Static Gaze' the robot's head is immobile and directed towards an intermediate position between tutor and table. To test the claim that tutors adjust their action presentation to a robot's online feedback, we wanted a control condition that would allow us to explore what happened if a robot did not produce any online feedback (as is the case with most existing Social Learning HRI studies). Our data set presents one case of a robot's static gaze (VP27). Its implications for the tutor's conduct will be examined here.

The semi-automatic tracking of the frog's motion reveals a set of even arc-shaped trajectories similar to those cases where the tutor did not orient to the robot (see section 5.2.1). In contrast, however, here the tutor does indeed orient to the robot.

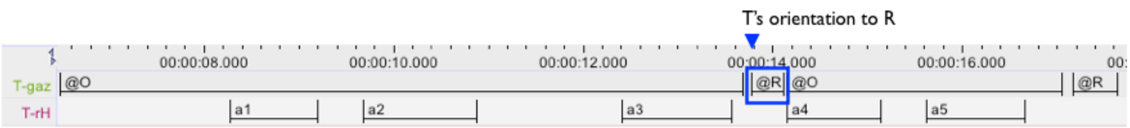


VP27 - 1st presentation   VP27 - 2nd presentation   VP27 - 3rd presentation

Figure 5: Tutor's successive action presentation in the robot's static gaze condition

During the first action presentation, the tutor looks at the robot after the first two jumps and before he starts another series of two jumps.

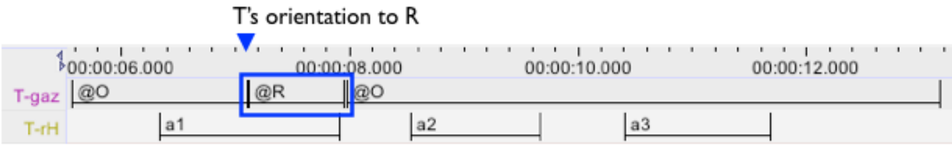
Fragment 8: VP27 – 1st presentation



Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

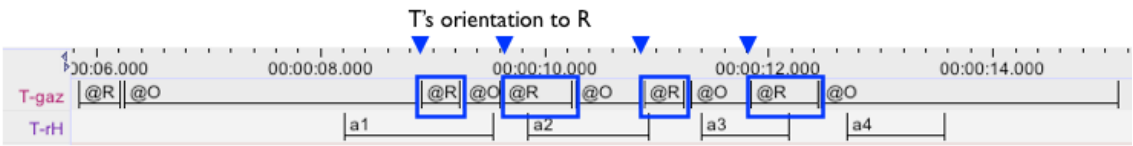
During the second presentation, the tutor gazes toward the robot beginning with his first jump.

Fragment 9: VP27 – 2nd presentation



During the third presentation, the tutor's gaze to the robot increases. He looks at it during the second half of the first jump, at the beginning and end of the second jump, and during the peak of the third jump.

Fragment 10: VP27 – 3rd presentation



In summary, this example shows that a robot's static gaze does not invite the tutor to adapt his action presentation either online or in the subsequent action presentation, despite being oriented toward the robot. Also, it seems that the tutors may be concentrated on performing the action during their first presentation, but then become more confident in their actions and more interested in the robot's conduct.

8. Summary and Implications

We began with the observation in HHI that tutors constantly monitor the recipient's reactions when presenting some action, and adjust the emerging action trajectory to their ongoing feedback (here: gaze; Pitsch et al. 2009, submitted). We used this as

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

motivation to model a robot's feedback behavior in a Social Learning scenario. In conducting an HRI study, we wanted to explore whether, and if so, how a robot's online feedback through gaze could pro-actively shape the way in which tutors performed action presentations. An analysis of 9 cases in which tutors presented how a toy frog jumps across a table, produced the following results:

(1) Human participants interpret a robot's conduct as being senseful and ascribe intentionality to the observed actions. The technical object 'robot' is thus conceived of as an actor who exerts its agency upon the world (Latour 1988).

(2) A robot, when using adequate online feedback strategies, can pro-actively shape the tutor's presentation, or more generally, his actions and conduct. These results draw attention to human and robot dyads as interactional systems where human adaptability to co-participants and changes in the environment is the most important resource. Our analysis revealed that tutors adjust the pauses between their actions, and the speed and height of their motions to the robot's shifting visual focus of attention. These are the central parameters described as "motionese" (Brand et al 2007, Nagai & Rohlfing 2009) features in adult-child-tutoring.

(3) Through their adjustments, tutors highlight specific aspects of the presented action and decompose it into sub-structures, making it visible as a phenomenon and helping it stand out from the background of the general interactional flow. This involves the following aspects: (a) sequencing of actions and building interactional units, (b) speed and rhythm, and (c) amplitude of an action.

(4) Note that the robot can also provoke *disturbances* in the tutor's performance (e.g. to engender a re-doing of parts of an ongoing action). As these require particular repair strategies and thus constitute an additional task for the robot, the system's behavior

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

would best be designed to allow for a tutor's disturbance-free presentation. The concrete interactional devices used by the robot as described in section 5 to 7 relate to the 'frog jump' action, so that their generalizability to other types of actions should be empirically tested.

(5) The form of the tutor's adaptation depends on his awareness of the robot's behavior. A robot thus needs strategies for organizing the human's focus of attention (Kuzuoka et al. 2008). By observing the tutor's gaze conduct, the system could also develop hypotheses about (un-)likely types of adaptation.

(6) Differences in the tutor's adaptation could be observed depending on the robot's different gaze conditions. For the robot's static gaze, no tutor adaptations were found. For the robot's action-related gaze, the tutor sequences his actions and adjusts the emerging action trajectories on a micro-level in step with the robot's behavior. In the random condition, the tutor also adapts on the level of sequence structures, but is otherwise more permissive with the robot's conduct; he does not expect the robot to precisely follow his hand motion. On the one hand, these observations on the micro-adjustment of trajectories are in line with the quantitative results from this study reported in Vollmer (2011), which suggest that actions are presented more slowly in the action-related (termed 'social gaze') than in the static gaze condition. On the other hand, the question arises as to how precise a robot's conduct needs to be to engage in successful interaction with a human (see also Pitsch et al 2012).

(7) In this study, only the first step of imitation learning was considered, namely, the tutor's action presentation. As in our data set, the robot mainly reproduces the action as a pre-programmed goal-based reproduction (emulation), it is not possible to study the effect of the robot's feedback on its actual learning. However, the data yields an

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

interesting observation. In two cases (VP18 and VP51) the tutor repeats his presentation particularly often (7 and 6 times as opposed to the average 3.8 times). These cases happen to also be the only ones where the robot's reproduction changes over time and appears to adapt (for different reasons). Comparing these observations with the quantitative results presented in Vollmer (2011), two different principles come to light. Vollmer (2011) shows that emulated actions were demonstrated more often than imitations, suggesting that the robot's failure to reproduce an action incites the tutor to continue the presentation. Our qualitative analysis rather points to the idea that changes in the system's conduct could motivate a tutor to continue the presentation.

In sum, the present study advances our understanding of robotic "Social Learning" in that it (i) suggests a paradigm shift towards considering human and robot as *one* interactional learning system and (ii) demonstrates how the robot can shape the tutor's actions through its online feedback. In the future, such analyses and ideas need to be considered in conjunction with turn-based feedback (Vollmer et al. 2010), and should be systematically integrated with sophisticated learning algorithms (Kim et al 2009) where the connection between the robot's feedback and actual progress realized as internal changes in the system can be investigated.

9. References

Alissandrakis, A., Syrdal, D. S., & Miyake, Y. (2011). Helping robots imitate: Acknowledgement of, and adaptation to, the robot's feedback to a human task demonstration. In K. Dautenhahn & J. Saunders (Ed.), *New Frontiers in Human-Robot Interaction* (pp. 9-33). Amsterdam/Philadelphia: Benjamins.

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

Brand, R. J., Shallcross, W. L., Sabatos, M. G., & Massie, K. P. (2007). Fine-Grained analysis of motionese: Eye gaze, object exchanges, and action units in infant-versus adult-directed action. *Infancy, 11(2)*, 203-214.

Breazeal, C., & Scassellati, B. (2002). Challenges in building robots that imitate people. In K. Dautenhahn & C. L. Nehaniv (Eds.), *Imitation in animals and artifacts* (pp. 363-389). Cambridge, MA: MIT Press.

Bruner, J. S. (1985). The role of interaction formats in language acquisition. In J.P. Forgas (Ed.), *Language and social situations* (pp. 31-46). New York: Springer.

Cangelosi, A., et al. (2010). Integration of action and language knowledge: A roadmap for developmental robotics. *Transactions on Autonomous Mental Development, 2(3)*, 167-195.

Dausendschön-Gay, U. (2003). Producing and learning to produce utterances in social interaction. *Eurosla Yearbook, 3*, 207-228.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27(2)*, 209-21.

Fogel, A. (1993). *Developing through relationships: Origins of communication, self, and culture*. Chicago, IL: University of Chicago Press.

Gergely, G., & Csibra, G. (2005). The social construction of the cultural mind: Imitative learning as a mechanism of human pedagogy. *Interaction Studies, 6(3)*, 463-481.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32(10)*, 1489-1522.

Herberg, J. S., Saylor, M. M., Ratanaswasd, P., Levin, D. T., & Wilkes, D. M. (2008). Audience-Contingent variation in action demonstrations for humans and computers. *Cognitive Science, 32(6)*, 1003-1020.

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

Hutchby, I. (2001). *Conversation and technology: From the telephone to the internet*. Malden, MA: Polity Press & Blackwell.

Kim, E. S., Leyzberg, D., Tsui, K. M., & Scassellati, B. (2009). How people talk when teaching a robot. In *Proceedings HRI 2009* (pp. 23-30).

Kuzuoka, H., Pitsch, K., Suzuki, Y., Kawaguchi, I., Yamazaki, K., Kuno, Yamazaki, A, Heath, C, & Luff, P. (2008). Effects of restarts and pauses on achieving a state of mutual gaze between a human and a robot. In *Proceedings CSCW 2008* (pp. 201-204).

Latour, B. (1988). Mixing humans and nonhumans together: The sociology of a door-closer. *Social Problems, 35(3)*, 298-310.

Mondada, L. (2006). Participants' online analysis and multimodal practices: Projecting the end of the turn and closing of the sequence. *Discourse Studies, 8(1)*, 117-129.

Mondada, L., & Pekarek-Doehler, S. (2000). Interaction sociale et cognition située. Quels modèles pour la recherche sur l'acquisition des langues? *Acquisition et Interaction en Langue Etrangère, 12*, 147-174.

Mühlig, M., Gienger, M., Hellbach, S., Steil, J. J., & Goerick, C. (2009). Task-level imitation learning using variance-based movement optimization. In *Proceedings ICRA 2009* (pp.1177-1184).

Nagai, Y., & Rohlfing, K. (2009). Computational analysis of motionese. Toward scaffolding robot action learning. *Transactions on Autonomous Mental Development, 10(1)*, 44-54.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences, 13(3)*, 423-451.

Pitsch, K., Vollmer, A. -L., Fritsch, J., Wrede, B., Rohlfing, K., & Sagerer, G. (2009). On the loop of action modification and the recipient's gaze in adult-child-interaction. In *Proceedings GESPIN 2009* (6 pages).

Pitsch, K., Vollmer, A.-L., Rohlfing, K., Fritsch, J., & Wrede, B. (submitted): Tutoring in adult-child-interaction: On the loop of action modification and the recipient's gaze.

Pitsch, K., Lohan, K.S., Rohlfing, K., Saunders, J., Nehaniv, C.L., & Wrede, B. (2012): Better be reactive at the beginning. Implications of the first seconds of an encounter for the tutoring style in human-robot-interaction. *Proceedings Ro-Man 2012* (pp. 974-981).

Steels, L., & Kaplan, F. (2002). Aibos first words: The social learning of language and meaning. *Evolution of Communication, 4(1)*, 3-32.

Suchman, L. (1987). *Plans and situated actions. The problem of human machine communication*. Cambridge: Cambridge University Press.

Vollmer, A. -L., Lohan, K. S., Fischer, K., Nagai, Y., Pitsch, K., Fritsch, J., Rohlfing, K., & Wrede, B. (2009). People modify their tutoring behavior in robot-directed interaction for action learning. In *Proceedings ICDL 2009* (pp. 1-6).

Vollmer, A. -L., Pitsch, K., Lohan, K., Fritsch, J., Rohlfing, K., & Wrede, B. (2010). Developing feedback: How children of different age contribute to a tutoring interaction with adults. In *Proceedings ICDL 2010* (pp. 76-81).

Vollmer, A.-L. (2011). *Measurement and analysis of interactive behavior in tutoring action with children and robots*. PhD thesis, Technical Deparment, Bielefeld University.

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge MA: Havard University Press.

Wertsch, J. V. (1985). *Vygotsky and the social formation of mind*. Cambridge, MA:

Final version published in Interaction Studies 14/2 (2013).
The final publication is available at https://benjamins.com/#catalog/journals/is/main

Havard University Press.

Wrede, B., Rohlfing, K., Hanheide, M., & Sagerer, G. (2008). Towards learning by interacting. In *Proceedings HRI 2008*, 139-150.

Zukow-Goldring, P., & Arbib, M. A. (2007). Affordances, effectivities, and assisted imitation: Caregivers and the directing of attention. *Neurocomputing, 70(13-15)*, 2181-2193.