

I Trained this Robot: The Impact of Pre-Experience and Execution Behavior on Robot Teachers*

Susanne Stadler¹ and Astrid Weiss² and Manfred Tscheligi¹

Abstract—The teacher–learner constellation is a special one in Human-Robot Interaction (HRI), as it can essentially improve intuitive interaction with robots. In a 2 (*background*: programmer vs. non-programmer) x 3 (*teacher*: self vs. believed other vs. other) between participants experiment (n=48, counter-balanced in gender), participants kinesthetically taught a humanoid NAO robot a specific behavior, which the robot had to execute afterwards. Next, participants downloaded a taught behavior to the NAO and were told that, the executed behavior either is (1) the one they previously taught (self), (2) the one someone else taught, but actually it was their own (believed other), or (3) the one someone else taught (other). We were interested in two main aspects: (1) whether programmers and non-programmers show differences in their teaching behavior and the perception of the teaching workload and (2) whether participants show a greater self-extension and trust into a robot they taught themselves over a robot they believed someone else taught. The study revealed that the teaching style independently of the *background* extends in the behavior execution time. Programmers showed a higher perceived workload than non-programmers. Differences in trust could not be found, but a self-extension effect was observed that people showed greater self-extension into a robot they taught themselves. Implications for Human-Robot Interaction are discussed.

I. INTRODUCTION

Human-Robot Interaction research is focusing a lot on questions concerning “natural and intuitive” interaction paradigms for naive users and how teaching and training of robots could look like in the future [1]. The introduction of robots into actual working and living environments will lead to the situation that users with different skills or *backgrounds* have to be able to teach robots different behaviors, which robots can execute autonomously afterwards. The approach of robot Programming-by-Demonstration (PbD) has always claimed to be an intuitive mean for teaching and interacting with robots [2] [3] and is well studied in terms of usability aspects such as efficiency [4] [5]. However, little research has been done so far to better understand how users with different backgrounds experience PbD not only in the teaching situation, but also in the situation when the robot executes a previously taught behavior. We stress the importance of perceived workload and programming pre-experience when

teaching a robot, as well as perceived self-extension and trust into the previously taught executed behavior to attain the user’s willingness to perform PbD tasks with robots in everyday life.

Consider a robot in a working context, e.g. a factory scenario, in which operators with different skill levels should be able to teach the robot various tasks and the robot should execute them appropriately after teaching. The teaching process should not cause a high workload for the operator and the operator should afterwards trust in the executed behavior. Moreover, in the working context it is highly likely that operators will not always train robots themselves, but also use behaviors co-workers previously have taught. Will this make a difference in system trust? Will operators trust significantly more in robot behaviors they taught compared to the ones others taught?

Other research suggests that people tend to perceive a robot as a part of themselves in some situations; in other words they feel an extension of their self into the robot [6]. As such we posit that PbD for robots alters the way how people experience the robot after it executes a previously taught task. Similarly, people feel differently about things they created themselves compared to things someone else created [7].

The current study explores two important aspects of PbD for Human-Robot Interaction, namely (1) whether programmers and non-programmers show differences in their teaching behavior and the perception of the teaching workload and (2) whether participants show a greater trust and self-extension into a robot they taught themselves over a robot they believed someone else taught.

The work is presented as follows: First related research on PbD in robotics is reviewed, also considering studies on teaching behavior, trust, and self-extension in HRI. Second, the study design including hypotheses, research methodology, and procedure is described, followed by the results of the user study. The last part of the presented work is dedicated to a discussion of the implications for HRI.

II. RELATED WORK

PbD appeared in the early 1980’s and made its way from the first ideas in software development up to the usage in robotics, especially in the area of manufacturing. In software development it was an intuitive and flexible way for users to program a computer simulation without having to learn a computer language [8]. Soon, researchers recognized the potential of PbD as an alternative to the tedious manual programming of robots as well as a way to achieve cost

*We greatly acknowledge the financial support by the Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development (Christian Doppler Laboratory for “Contextual Interfaces”) and from the Austrian Science Foundation (FWF) under grant agreement T623-N23, V4HRC.

S. Stadler and M. Tscheligi are with the ICT&S Center, University of Salzburg, 5020 Salzburg, Austria firstname.lastname@sbg.ac.at

A. Weiss is with the ACIN Institute of Automation and Control, Vienna University of Technology, 1040 Vienna, Austria astrid.weiss@tuwien.ac.at

reduction of the development and maintenance of robots in a factory. A first approach in robotics was symbolic reasoning [9]. Early studies were made to explore processes within this approach such as teach-in, guiding, and play-back methods (manual (teleoperated) control).

Around the early 1990's research appeared in the direction of Machine Learning (ML) techniques to extend the direct repetition process prevailing until then with record and play to one of generalization. This allows to transfer the generated programs to new robots or to products with different variations. The commonly used ML technique in Robot Programming-by-Demonstration (RPbD) is to learn a policy from examples provided by a teacher and after training, the robot recodes and reproduces the demonstrated actions. A survey of RPbD and its diversity of types can be found in [1]. This is in contrast to other techniques, e.g. reinforcement learning in which a policy is obtained by exploration. RPbD is considered as an intuitive communication medium for humans and is appropriate for experts as well as non-experts of robotics [5].

For a positively experienced RPbD in HRI both aspects are relevant: the user teaching situation and the perception of the robot behavior execution. Thus, subsequently we have a look on related research on human teaching behavior (section II-A) and on the two factors self-extension and trust in the perception of robot behavior (section II-B).

A. Human-Teaching Behavior

Little research has been conducted so far on the actual human teaching behavior in HRI. Thomaz and Breazeal [10] stress that people develop a mental model of the learner; they use the reward channel for both guidance and feedback. Furthermore, they require a separate channel for motivational feedback. Investigating the characteristics a robot needs for efficient work with a human partner, Breazeal et. al [4] refer to joint intention and learning theory [11], [12], [13]. They assume that humans are used to teaching in a social and collaborative way and usually choose the dialog as a form for tutelage. Grice [13] emphasizes the importance of confirmations for mutual agreement on accomplishment of a task and highlight the importance of making the collaborative work with robots not only efficient but also enjoyable for humans by using social skills [4].

B. Trust and Self-Extension in HRI

There is no doubt that trust plays an important role and impacts the effectiveness of human-human as well as human-robot cooperation. Luhmann [14] defined that "Interpersonal trust is a vital component of personal and work relationships as trust provides the foundations for accepting the unknown and coping with complexities that are beyond our control, leading to new opportunities to acquire new knowledge and experience." Regarding trust in automation Lee and Moray [15] defined that trust is "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability."

McKnight et al. [16] presented in their article definitions and measures towards trust and stress that trust is a primary predictor of technology usage. Billings et al. [17] raised the question if human-interpersonal trust is a good analogy for human-robot trust and additionally introduced trust in human-robot teams. They claim that among the most important robot characteristics for trust are performance, appearance and proximity. Further they noted that manipulating these design aspects and reliability of the robot can support the calibration of trust to an appropriate level for facilitating successful interaction.

We carry this thought one step further, considering that trust is also calibrated by the fact who manipulated the behavior of the robot: me or someone else as this might impact the experience of human-robot teams.

Similarly, we are interested in the concept of self-extension and if it carries over to RPbD in HRI. A person is particularly likely to extend one's sense of self into objects that one controls, creates, or personalizes [6]. Groom et al. [6] evaluated in a study how assembling and operating a robot as well as the form of the robot influences the degree of self-extension to the robot. Results show greater self-extension into a robot and preferred the personality of the robot people assembled over a robot they believed to be assembled by another. People who assembled the robot they operated showed greater overlap in terms of personality with the robot, were more attached to it, and felt more like a team. They also perceived the robot to be more like themselves, because they had more familiarity with the inner working of the robot. Groom et al. [6] also found, that human-like physical attributes minimize self-extension, which was also noted by Kiesler and Kiesler [18]. The reason for this effect is that an anthropomorphic robot is perceived as unique identity.

We assume that this aspect of self-extension should also appear when a user observes the behavior execution of a robot that was previously kinesthetically taught.

III. STUDY DESIGN

We used a 2 (*background*: programmer vs. non-programmer) x 3 (*teacher*: self vs. believed other vs. other) between participants experiment design. All participants had to teach the NAO robot a specific "pick-and-place" task. This type of task is commonly considered in robotic research for the factory context and explained in more detail in section III-B below. First the participants had to teach NAO with kinesthetic teaching ("Teaching Phase") and then they had to download ("Behavior Transfer") a previously taught behavior to the robot. We were interested if the pre-experience in programming affects the kinesthetic teaching and if people do perceive a robot differently depending on what they are told who has trained the robot: (1) they themselves (self), (2) someone else, but actually it was their own taught behavior (believed other), or (3) someone else (other).

We anticipated that people with programming experience need less demonstrations and experience a lower workload while teaching the robot. Furthermore, we expected people

to trust and self-extend more into a robot they have trained themselves.

These assumptions led to the following four research hypotheses:

- H1. People with programming experience will need less demonstrations to teach the robot.
- H2. People with programming experience will experience a lower workload while training the robot.
- H3. People will self-extend more into a robot they taught themselves than someone else.
- H4. People will trust a robot they taught themselves more than someone else.

A. Participants

In total 48 people aged between 18 and 55 years (mean=29.25, SD=5.13) participated in the study. The selection was gender-balanced within all conditions (*background, teacher*). We have chosen the age range 18 to 55 because this is a good representation of employees who might come into touch with robot programming in the work context. As mentioned before the condition “*background*” only differs within the degree of programming experience. People with programming experience in our study are either professional programmers or computer science students. In contrast people without programming experience never engaged with any kind of programming, neither in their everyday lives nor for professional reasons. All participants were German native speakers and the study was conducted in German language.

B. Study Setup

The study was set up as a factory scenario in which the user should teach a robot to pick up a box and place it in a machine. For this purpose we built a mock-up factory with two different machines (M1 and M2, see Fig. 1). The humanoid robot NAO served as a platform for the implementation of *kinesthetic teaching*, the selected RPbD approach.

During the “Teaching Phase” the participant taught the robot a complete behavior sequence to pickup a box, carry it to a machine and put it in there.

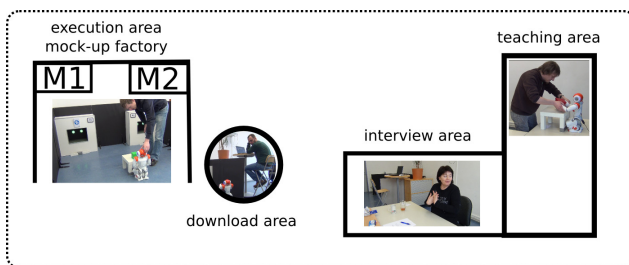


Fig. 1. Floor plan of the study setup: left execution area for the robot to perform, middle download area “Transfer Behavior”, left teaching area “Kinesthetic Teaching”

Within the “Behavior Transfer” phase the task of the participant was to download a behavior via a web-interface to the robot. Therefore, we implemented a web-interface to give the participants the possibility to download a behavior

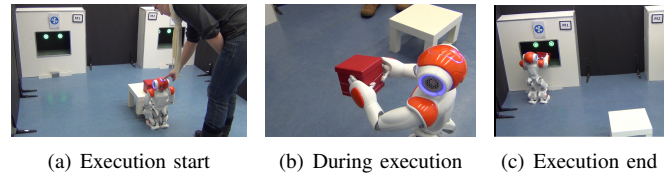


Fig. 2. The NAO robot within the execution area

sequence for picking up and placing a box in the machine, which the robot should then execute. This was a very simple web-interface offering three different saved behaviors including the self-taught behavior as well as two behaviors provided by other persons. For each training data set a little description was shown together with the teacher’s name and a choice of the target machine. In the following section we outline the study procedure in more detail.



Fig. 3. “Behavior Transfer” via web-interface

C. Procedure

In a first step participants were welcomed and asked to fill in two questionnaires besides a data usage consent form. One questionnaire collected demographic data, the other gathered information about existing programming and robotic experience. Robotic experience was not counter-balanced.

In phase₁ named “Kinesthetic Teaching” the participant had to teach the robot to pick-and-place a box. During phase₂ “Behavior Transfer” the participant’s task was to transfer an existing behavior to the robot by downloading the behavior via the web-interface. All participants first run through phase₁ and afterwards phase₂. Figure 4 outlines the procedure within the phases which we describe in detail below.

a) “*Kinesthetic Teaching*” (phase₁): *Teaching and execution of taught behavior*: The mandatory *restricted teaching* and the *optional teaching* comprise the kinesthetic teaching parts. During each teaching the participant is guiding the robot to pick-and-place a box. After each teaching performance the robot reproduces the taught behavior as a feedback for the participant. Subsequently, the user could choose to repeat the teaching of the robot if he or she was not satisfied. During the *restricted teaching* the participants had at most four tries to be satisfied with the teaching. This restriction was set to increase the workload level for the

participants and simulate a programming task which has to be completed with limited resources.

Afterwards the user started the execution by pressing the start button on the head of the robot (see Fig. 2a.). The robot picked up the box (as taught by the user), walked autonomously to the machine (see Fig. 2b.), and placed the box (also taught by the user) there (see Fig. 2c.). Now the participant saw for the first time if the previously taught behavior was satisfying within the environment of the mock-up factory. If the requirements were not met (e.g. the robot lost the box during transport due to loose grip) the participants were allowed to re-train the robot without any restriction on the number of trails. This stage is represented by *optional teaching*, cf. Fig. 4.

Questionnaires and interview: To investigate aspects of self-extension, knowledge about the participants' personality is needed. Fig. 4 shows the time-line of the study procedure which took up to two hours and introduces abbreviations. The participants had to fill in a personality questionnaire three times following Jacobs and Scholl [19]: with QP₁ personality traits of the participant were determined, whereas with QR₁ and QR₂ the participants had to complete the same questionnaires for the robot. The Nasa-RTLX questionnaire (QN₁) investigated the participants' workload for programming the robot. With the last questionnaire within this phase we collected data concerning *trust* (QT₂). The participants completed this phase with an interview I₁ about the intuitiveness of the kinesthetic teaching approach.

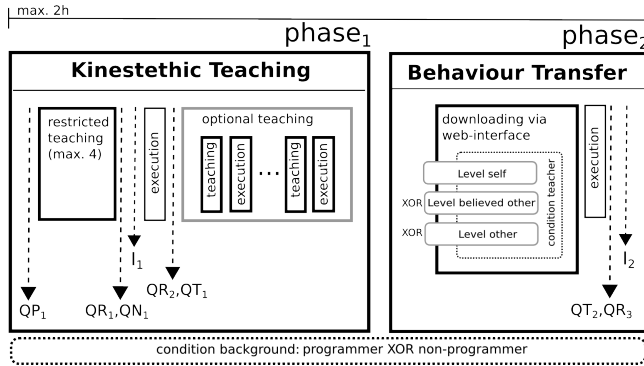


Fig. 4. Detailed overview of study phases: QP: Questionnaire participant's personality, QR: Questionnaire robot's personality, QN: Questionnaire Nasa-RTLX, QT: Questionnaire trust, I: Interview.

b) **"Behavior Transfer" (phase₂):** The participant's activity within this phase was to transfer a certain behavior to the robot. Each participant got either the assignment (i) "download the training data you taught to the robot", or (ii) "download the training data of another person to the robot". Thereby we wanted to evaluate if people tend to self-extend and trust more into a robot they taught themselves vs. a robot trained by someone else. For the condition "teacher" we defined three levels and assigned 16 participants to each (gender and background balanced within the levels).

Level self: The participant got assignment (i) and the behavior downloaded to the robot was indeed the behavior previously taught by the participant during phase₁.

Level believed other: The user was misled: Although assigned (ii), the behavior downloaded to the robot was not another person's training data, but in fact his or her own. This level was intentionally designed to control for the effect that participants cannot distinguish a self-taught behavior from a behavior someone else taught. We assume that a self-extension effect should appear also at this level, as participants should rate the behavior execution of the robot and not what they were told.

Level other: The participant got assignment (ii) and the program was really trained by another person.

After downloading the robot executed the transportation of the box within the execution area, according to the behavior transferred. This was followed by an other round of questionnaires: trust (QT₂) and personality (QR₃). QR₃ is targeted at the robot, not the participant. At the end a short interview I₂ is conducted about the overall acceptance followed by the debriefing.

D. Measures

1) **Teaching Observations:** We count the number of teachings and record the time each teaching takes. At first participants were told to have at most four repetitions to teach the "pick-and-place" task. Later on participants were allowed to optionally re-train the robot.

The *user-teaching-time* (overall time of a demonstration) can be divided into the *robot-idle-time* and the *robot-execution-time*. *Robot-idle-time* is the time the robot does nothing, because participants were shy, or afraid to move the robot's arms, or they were planning or thinking how to fulfill the task with the robot. The *robot-execution-time* is the period the robot was in motion to execute the previously taught or downloaded pick-and-place task.

2) **Teaching Workload:** We used the Nasa-RTLX questionnaire (QN₁) to evaluate the workload of the participants immediately after the *restricted training*.

3) **Self-Extension:** We asked the participants to fill in the personality questionnaire from Jacobs and Scholl [19] four times: initially to assess the personality of the participant (QP₁) and subsequently three times to assess the personality of the robot after restricted training (QR₁), optional training (QR₂), and download (QR₃).

The first three measurements permit to check whether the participant rates the personality of the robot after kinesthetic teaching in a similar way compared to his or her own personality. The last measurement rates the personality of the robot after download. We measured PA (self-assured, dominant) and HI (subservient, shy). Participants rated how well thirty words, such as "shy", "confident" or "inconsiderate" described them. They indicated their responses on eight-point scales ranging from "Strongly disagree" to "Strongly agree".

4) *Trust*: In order to investigate trust and based on the lack of available trust questionnaires for our requirements, we composed a questionnaire consisting of selected dimensions from existing questionnaires. To measure *functionality* and *reliability* as dimensions of trust, we took the according items from McKnight's questionnaire [16]. Additionally, *System Trust* was measured by applying the *System Trust* items defined by Master [20]. *Perceived Usefulness* and *Perceived Ease of Use* (as dimensions of technology acceptance) were taken from Venkathesh and Davis [21]. The rating was on a 5-point Likert scale (1-Strongly disagree to 5-Strongly agree). *Functionality* refers to whether one expects a technology to have the capacity or capability to complete a required task [16]. *Reliability* suggests one expects a technology to work consistently and predictably [16]. *Perceived Usefulness* is defined as the degree to which a user believes that using the system will enhance his or her performance. *Perceived Ease of Use* is defined as the degree to which the user believes that using the system will be free from effort. All scales summatively give an overall system trust rating.

IV. RESULTS

A. Teaching Behavior

The results are divided into two aspects. The first deals with a) the number of demonstrations and the second b) focuses on the time per demonstration.

a) In total 48 participants trained the robot 218 times. On average the number of demonstrations per participant was 4.54 (SD=2.35). The training phase was divided into a *restricted teaching* (max. four times) and an *optional teaching* (no restrictions on the number of demonstrations). During the *restricted teaching*, all participants (n=48) trained the robot 116 times (mean=2.42, SD=0.79). During the *optional teaching*, 35 participants trained the robot 102 times (mean=2.91, SD=2.25). 13 participants choose not to train again.

As mentioned before our first assumption was that programmers would performed fewer demonstrations. As Table I shows there is hardly any difference between the average number of demonstrations of non-programmers and programmers in the *overall teaching* as well as the *restricted teaching*. However, a difference becomes apparent in the *optional teaching*, showing that non-programmers tend to do more trainings. This difference was not statistically significant (Wilcoxon-rank sum test: $W = 199.5$, $p > .05$).

TABLE I

AVERAGE NUMBER OF DEMONSTRATIONS, OVERALL, RESTRICTED AND OPTIONAL TEACHING, P (PROGRAMMER), NP (NON-PROGRAMMER)

	average number of demonstrations		
	overall teaching	restricted teaching	optional teaching
P	4.37 $SD = 2.10$	2.42 $SD = 0.78$	2.47 $SD = 2.12$
NP	4.71 $SD = 2.61$	2.42 $SD = 0.83$	3.44 $SD = 2.37$

b) Similarly, comparing the teaching times in seconds between programmers and non-programmers did not result in significant differences, see Table II. However, there was a significant correlation between the *user-teaching-time* and the *robot-execution-time* of the taught behavior, $r = 0.87$, $p < .001$. Programmers were affected more than non-programmers, similarly male participants more than female participants (see Table II). This evidence suggests that the *user-teaching-time* is in a strong relationship with the *robot-execution-time*. The longer the *user-teaching-time* takes, the longer is the *robot-execution-time*. This might indicate that self-extension is even observable on a behavioral level.

TABLE II

CORRELATION (r , SPEARMAN, *** = $p < .001$), USER-TEACHING-TIME WITH ROBOT-EXECUTION-TIME. P (PROG.), NP (NON-PROG.)

	r	average user-teaching-time (s)	average robot-execution-time (s)
all users	0.87***	50.48 $SD = 29.07$	28.47 $SD = 16.32$
P	0.92***	48.01 $SD = 24.51$	28.78 $SD = 16.17$
NP	0.86***	52.78 $SD = 32.68$	28.19 $SD = 16.53$
male	0.93***	51.72 $SD = 28.28$	30.57 $SD = 16.67$
female	0.64***	49.17 $SD = 29.95$	26.26 $SD = 15.71$
P-male	0.98***	44.37 $SD = 21.15$	27.81 $SD = 13.15$
P-female	0.67***	52.67 $SD = 27.79$	30.02 $SD = 19.45$
NP-male	0.90***	59.91 $SD = 32.84$	33.64 $SD = 19.55$
NP-female	0.67***	46.48 $SD = 31.47$	23.37 $SD = 11.47$

B. Teaching Workload

We computed the Nasa-RTLX (QN₁) score to explore the perceived workload during teaching. The workload for programmers (mean=42.5, SD=11.64) in contrast to non-programmers (mean=36.84, SD=13.68) to teach the NAO robot did not differ significantly $W=237$, $p > .05$, $r = -0.15$ (Wilcoxon rank sum test). Consequently, we cannot confirm our hypothesis that programmers experience a lower workload than non-programmers.

However, we found a significant difference when we additionally consider pre-experience with robots. On average, programmers with robot experience perceived a greater workload (mean=44.79, SD=13.22) than non-programmers with robot experience (mean=33.06, SD=11.96). This difference was significant $t(21.78) = 2.28$, $p < .05$ and it did represent a middle-sized effect $r = 0.439$ (t-test).

During the interviews we revealed that based on a better idea of the complexity of the robot and the underlying procedures, programmers with robot experience are concerned about missing information on error handling. Therefore they try to perform the teaching very diligently which causes a perceived higher workload. The majority of the programmers with robot experience stated in the interviews that they consider the kinesthetic approach as an intuitive way to program the robot. They also all noted that in order to handle the robot in an appropriate way they need more experience with it: "It requires experience in the use". In contrast, non-programmers with robot experience seem to be completely unbiased.

C. Self-Extension

We measured four times the personality QP_1, QR_1, QR_2, QR_3 (see subsection III-D.3 and Fig. 4) to evaluate if people tend more to self-extend in a robot they taught themselves than in a robot taught by another person.

When participants taught the robot themselves and downloaded the behavior knowingly to the robot, we found no significant difference between the QP_1 and the QR_3 ratings (*Level self*, $T(153) = 1.13$, $p > .05$). In contrast, when participants believed the downloaded behavior was from another person the rating differs significantly, no matter if the behavior really was taught by themselves (*Level believed other*, $T(153) = 3.64$, $p < .001$) or by another person (*Level other*, $T(153) = 2.06$, $p < .04$).

TABLE III

AVERAGE SCORE: SEMANTICAL DIFFERENTIAL (1=HI (SUBSERVIENT, SHY) TO 8=PA (SELF-ASSURED, DOMINANT))

	<i>Level</i>		
	<i>self</i>	<i>believed other</i>	<i>other</i>
PA/HI (QP_1)	4.83 SD=0.14	5.09 SD=0.13	4.84 SD=0.15
PA/HI (QR_1)	4.44 SD=0.13	4.78 SD=0.14	4.80 SD=0.13
PA/HI (QR_2)	4.40 SD=0.13	4.57 SD=0.13	4.67 SD=0.14
PA/HI (QR_3)	4.58 SD=0.14	4.50 SD=0.13	4.48 SD=0.14

In other words, the data supports that participants in the self level demonstrated a greater overlap in personality ratings, than in the other two levels. However, our hypothesis is only partly supported, as a self-extension effect should have also happened at the level where the robot executes the self-taught behavior, even though the user is misled and is told that some else taught the robot. In other words it is more important for people's assessment of the executed behavior what they were told than the actual behavior execution. Interestingly, this effect only holds true for participants with programming experience. For non-programmers no significant differences could be observed at all, which indicates that a feeling of self-extension is always present for naive users.

D. Trust

As mentioned before we measure *Functionality* and *Reliability* as dimensions of trust, *Perceived Ease of Use* and *Perceived Usefulness* as dimensions of technology acceptance and *System Trust*. We assessed the degree of trust using QT_1 and QT_2 after "Kinesthetic Teaching" and "Behavior Transfer", respectively.

We found no significant differences between the trust ratings (QT_2) for *Level self*, *Level believed other*, *Level other*, no matter for overall trust or the sub-dimensions. As a consequence H4 has to be rejected. Also we compared the trust ratings (QT_1 and QT_2), without finding a significant difference; see Table IV. The sub-scale *Perceived Ease of Use* (PEU) is not significant by a very small margin $T(29) = -2.04$ $p = .0502$.

TABLE IV

AVERAGE TRUST SCORE, T=TRUST, ST=SYSTEM TRUST, TA=TECHNOLOGY ACCEPTANCE, R=RELIABILITY, F=FUNCTIONALITY, PU=PERCEIVED USEFULNESS, PEU=PERCEIVED EASE OF USE

	QT_1	QT_2	yuend/t-test
overall	3.50 SD= 0.63	3.23 SD=1.03	$t = -0.28$ $df = 93.83$ $p > .05$
T	3.35 SD=0.83	3.32 SD=0.78	$T(29) = 0.55$ $p > .05$
ST	3.53 SD=0.60	3.55 SD= 0.63	$T(29) = 0.04$ $p > .05$
TA	3.46 SD=0.83	3.60 SD=0.74	$T(29) = -1.18$ $p > 0.5$
R	3.12 SD= 0.85	3.13 SD=0.80	$T(29) = 0.44$ $p > .05$
F	3.60 SD= 0.97	3.62 SD=0.97	$T(29) = -0.45$ $p > .05$
PU	3.17 SD=1.22	3.20 SD=1.18	$T(29) = -0.17$ $p > .05$
PEU	3.76 SD=0.90	4.00 SD=0.70	$T(29) = -2.04$ $p < .06$

V. DISCUSSION

A. Summary and interpretation of results

Not all of our hypotheses were supported by data. The predicted effect in H1 that people with programming experience will need less demonstrations to teach the robot could not be supported. However, we could observe another interesting effect, namely that longer *user-teaching-time* relates to a longer *robot-execution-time*. This shows that there is a teaching behavior transfer which actually affects the robot behavior execution. This indicates a self-extension effect not only on the attitudinal level, but also on the behavioral one.

Similarly, H2 was not supported by data in the way we expected it. No significant differences regarding perceived workload (measured with the Nasa-RTLX questionnaire) could be identified. However, another interesting effect could be found in the data: programmers with robot pre-experience perceived a significantly higher workload compared to non-programmers with robot pre-experience. We assume that programmers with robot experience are more sensitized to handle the robot correctly and are more aware of potential errors and therefore experience a higher workload.

As predicted by H3, participants in *level self* demonstrated a greater overlap in personality ratings than participants in the other levels where they were told the robot was trained by someone else (*believed other* and *other*). Moreover, we found that the programming pre-experience impacts the self-extension effect: non-programmers always show a self-extension effect, whereby participants with experience seem to be misled by what the experimenter tells them about the robot behavior and despite that information assess every robot behavior equally.

These findings provide further evidence that PbD causes a perceived extension of the self for naive users, however as the other results showed there is no effect on *System Trust*. In other words H4 that participants will trust a robot they taught themselves more has to be rejected.

B. Implications for HRI

These findings have implications for the user's willingness to perform PbD tasks with robots in everyday life. For example, if we go back to the working context scenario, the results support that it does not matter who program a robot in terms of *System Trust*, even though there is a higher self-extension, but only for operators without any

pre-experience in programming. On the other hand, in a domestic or care context application a higher self-extension can positively influence the human-robot relationship and increase the bonding to personal service robots above all for naive users.

Similarly, our study supports the claim that PbD is an intuitive mean for naive users to train robots, but that pre-experience in robotics and programming increases the perceived workload as users are more careful and reflective about how to train the robot. This might have an impact on using PbD in the industrial context (e.g. the Baxter robot [22]) and strategies for error handling and maintenance by experienced operators should be considered in future.

C. Limitations

There are several limitations to this study. First our study was conducted in a factory mock-up and the artificial and very controlled study setting might have had an impact on the trust and self-extension ratings. Second the usage of the NAO robot which only performed one task, which a human could have performed faster in the study setup, might have impacted the self-reporting measures as well. Future studies should be done with different robots and tasks not only for a working scenario, but also domestic and care scenarios as this might produce different results. Clearly long-term studies on how the teaching behavior and the perception of the robot execution behavior are also required to validate our findings.

VI. CONCLUSION

Programming-by-Demonstration is considered as intuitive mean to teach and interact with robots, however studies in this field tend to be limited on efficiency aspects. This study indicates that the teacher experience differs depending on the pre-experience and that the perception of the executed behavior differs depending on the belief who taught the robot (self vs. other). These results suggest that for PbD in HRI not only the teaching phase needs to be considered but also the behavior execution phase to optimize for the users' willingness to teach robots that way in everyday life.

REFERENCES

- [1] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469–483, 2009.
- [2] C. L. Breazeal, "Sociable machines: Expressive social exchange between humans and robots," Ph.D. dissertation, Massachusetts Institute of Technology, 2000.
- [3] M. J. Matarić, "Behavior-based robotics as a tool for synthesis of artificial behavior and analysis of natural behavior," *Trends in Cognitive Sciences*, vol. 2, no. 3, pp. 82–86, 1998.
- [4] C. Breazeal, G. Hoffman, and A. Lockerd, "Teaching and working with robots as a collaboration," in *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 3*. IEEE Computer Society, 2004, pp. 1030–1037.
- [5] A. Weiss, J. Igelsbock, S. Calinon, A. Billard, and M. Tscheligi, "Teaching a humanoid: A user study on learning by demonstration with HOAP-3," in *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009. RO-MAN 2009*. IEEE, 2009, pp. 147–152.
- [6] V. Groom, L. Takayama, P. Ochi, and C. Nass, "I am my robot: The impact of robot-building and robot form on operators," in *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '09. New York, NY, USA: ACM, 2009, pp. 31–36. [Online]. Available: <http://doi.acm.org/10.1145/1514095.1514104>
- [7] R. Belk, *Possessions and Self*. Wiley Online Library, 1988.
- [8] A. Cypher, D. C. Halbert, D. Kurlander, H. Lieberman, D. Maulsby, B. A. Myers, and A. Turransky, *Watch What I Do: Programming by Demonstration*, 1st ed. The MIT Press, 1993.
- [9] T. Lozano-Pérez, "Robot programming," in *Proceedings of the IEEE*, vol. 71, Jul. 1983, pp. 821–841.
- [10] A. L. Thomaz and C. Breazeal, "Teachable robots: Understanding human teaching behavior to build more effective robot learners," *Artificial Intelligence*, vol. 172, no. 6, pp. 716–737, 2008.
- [11] L. L. S. Vygotskii, *Mind in society: The development of higher psychological processes*. Harvard university press, 1978.
- [12] P. R. Cohen, H. J. Levesque, J. H. Nunes, and S. L. Oviatt, *Task-oriented dialogue as a consequence of joint activity*. SRI International, 1990.
- [13] H. P. Grice, "Logic and conversation," *Syntax and Semantics*, vol. 3, pp. 41–58, 1975.
- [14] N. Luhmann, "Trust and power: Two works." 1979.
- [15] J. Lee and N. Moray, "Trust, control strategies and allocation of function in human-machine systems," *Ergonomics*, vol. 35, no. 10, pp. 1243–1270, 1992.
- [16] D. H. Mcknight, M. Carter, J. B. Thatcher, and P. F. Clay, "Trust in a specific technology: An investigation of its components and measures," *ACM Trans. Manage. Inf. Syst.*, vol. 2, no. 2, pp. 12:1–12:25, Jul. 2011. [Online]. Available: <http://doi.acm.org/10.1145/1985347.1985353>
- [17] D. R. Billings, K. E. Schaefer, J. Y. Chen, and P. A. Hancock, "Human-robot interaction: Developing trust in robots," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '12. New York, NY, USA: ACM, 2012, pp. 109–110. [Online]. Available: <http://doi.acm.org/10.1145/2157689.2157709>
- [18] T. Kiesler and S. Kiesler, "My pet rock and me: An experimental exploration of the self extension concept," *Advances in Consumer Research*, vol. 32, 2004.
- [19] I. Jacobs and W. Scholl, "Interpersonale Adjektivliste (IAL)," *Diagnostica*, vol. 51, no. 3, pp. 145–155, 2005.
- [20] R. Master, X. Jiang, M. T. Khasawneh, S. R. Bowling, L. Grimes, A. K. Gramopadhye, and B. J. Melloy, "Measurement of trust over time in hybrid inspection systems," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 15, no. 2, pp. 177–196, 2005. [Online]. Available: <http://dx.doi.org/10.1002/hfm.20021>
- [21] V. Venkatesh and F. D. Davis, "A theoretical extension of the technology acceptance model: Four longitudinal field studies," *Manage. Sci.*, vol. 46, no. 2, pp. 186–204, Feb. 2000. [Online]. Available: <http://dx.doi.org/10.1287/mnsc.46.2.186.11926>
- [22] E. Guizzo and E. Ackerman, "The Rise of the ROBOT WORKER," *IEEE Spectrum*, vol. 49, no. 10, pp. 34–41, Oct. 2012.