Teaching Robots Style: Designing and Evaluating Style-By-Demonstration for Interactive Robotic Locomotion

James E. Young[1], Ehud Sharlin[2], Takeo Igarashi[3,4]

[1]*University of Manitoba, Winnipeg, Canada*

[2]*University of Calgary, Calgary, Canada*

[3]*University of Tokyo, Tokyo, Japan*

[4]*JST ERATO, Tokyo, Japan*

In this paper we present a multi-part formal design and evaluation of the *style-by-demonstration* (SBD) approach to creating interactive robot behaviors: enabling people to design the *style* of interactive robot behaviors by providing an exemplar. We first introduce our Puppet Master SBD algorithm that enables the creation of interactive robot behaviors with a focus on style: users provide an example demonstration of human-robot interaction and Puppet Master uses this to generate real-time interactive robot output that matches the demonstrated style. We further designed and implemented original interfaces for demonstrating interactive robot style and for interacting with the resulting robot behaviors. Following, we detail a set of studies we performed to appraise users' reactions to and acceptance of the SBD interaction design approach, the effectiveness of the underlying Puppet Master algorithm, and the usability of the demonstration interfaces. Fundamentally, this paper investigates the broad questions of how people respond to SBD interaction, how they engage SBD interfaces, how SBD can be practically realized, and how the SBD approach to social human-robot interaction can be employed in future interaction design.

1. INTRODUCTION

Research in human-robot interaction (HRI) has shifted away from viewing robots primarily as industrial tools, toward considering how people interact with robots in everyday spaces [Forlizzi and DiSalvo, 2006, Kiesler and Hinds, 2004, Sung et al., 2009]. One recurring theme is that people have a tendency to anthropomorphize robots and treat them as social actors [Bartneck et al., 2007, Forlizzi and DiSalvo, 2006]; thus it is important to consider how human-centric interaction elements such as robot personality, behavior style, and emotion relate to human-robot interaction design. [Young et al., 2010c]

This paper deals with *style by demonstration* (SBD), a new view of the classic programming-by-demonstration (PBD) approach that refocuses onto and emphasizes the *style* of a robotic action rather than a task-oriented goal [Young et al., 2010a]. Instead of learning, for example, a particular navigation route, with SBD the robot learns the motion style it should use to traverse the route, such as moving aggressively or timidly. The robot can communicate such characteristics using the *style* of its movements, where *style* can be defined as a robot's "expressive movement…the way in which behavior is performed" [Gallaher, 1992]. Our SBD approach builds directly upon the ongoing success of programming-by-demonstration for robots (e.g., [Breazeal, 2002, Frei et al.,

CONTENTS

1. INTRODUCTION 1
2. RELATED WORK 3
3. SBD INTERFACES FOR STYLISTIC ROBOT LOCOMOTION 4
   3.1. The Characteristic Robot Sounds Extension 5
   3.2. Interaction-Environment Test Bed 6
   3.3. Demonstration With the *Broomstick* Interface 6
   3.4. Demonstration With the *Surface* Interface 7
4. THE ROBOTIC PUPPET MASTER ALGORITHM 8
5. STUDY APPROACH AND DESIGN 9
   5.1. *Programmer* Design Critique 11
   5.2. Current Studies: Qualitative Evaluation Approach 12
6. DEMONSTRATION STUDY 14
   6.1. Tasks and Procedure 14
   6.2. Quantitative Results 16
   6.3. Qualitative Results 16
   6.4. Discussion 19
7. OBSERVER STUDY 21
   7.1. Tasks and Procedures 22
   7.2. *Exploratory* Phase Results 23
   7.3. Matching Phase Results 24
   7.4. In-Situ Phase Results 24
   7.5. Post-Test and Overall 24
   7.6. Discussion 26
8. REFLECTIONS AND IMPLICATIONS 27
   8.1. Implications and Guidelines 28
9. CONCLUSIONS 28

2000, Matsui et al., 2005, Raffle et al., 2004]), both in terms of technically enabling robots to learn difficult tasks from demonstrations, and in making the teaching interface accessible to people with no robotics expertise: SBD leverages users' existing *social stock of knowledge* [Berger and Luckmann, 1966] understanding of teaching others through demonstration.

Such style-laden robotic motion could be easily achieved by simply repeating a recorded demonstration; a technique used for choreographed and pre-programed robot actions. What differentiates our SDB approach – and is a crucial component of both our interaction and study design – is that we target *interactive* robot behaviors: robots that convey personality by how they interact in real-time with a user, an unpredictable counterpart person. For example, a robot may interactively follow someone in a way that would be perceived as polite or reluctant, making appropriate reaction changes as the person alters their walk or route, or a robot could react to a person's presence and actions with movements that would be perceived as being happy or enthusiastic. This dynamic interactive behavior cannot be achieved by replaying a static action.

Our SBD interfaces, algorithm, and evaluations focus specifically on how a robot can communicate through the style of its locomotion path only – how it moves about a space to interact with a counterpart person – and on how the robot can learn such interaction styles from user-provided demonstrations.

A primary contribution of this paper is our evaluation that targets three key goals: 1) we explore how people respond to the core SBD concept of teaching robots to communicate using interactive locomotion style, 2) we test the usability of our particular demonstration and robot interaction interfaces, and 3) we test the efficacy of the underlying learning technique, our robotic Puppet Master SBD learning algorithm [Young et al., 2010b] – a system that has not previously been evaluated. Overall, we aim to develop grounded and broad understanding of how people engage SBD and the related demonstration interfaces, and of how they interact with robots that communicate using interactive, stylistic locomotion. We finish this paper by distilling our evaluation results as straightforward design guidelines and lessons learned that we believe will have implications for other SBD systems, other robots that communicate via interactive style, and for HRI in general.

2. RELATED WORK

People attribute style, personality and emotions to even simple abstract movements [Heider and Simmel, 1944], and thus we argue that style-related properties of robotic movement form a fundamental aspect of HRI that cannot be ignored: SBD explicitly leverages this communication channel. Researchers, particularly in animation, have developed methods for leveraging this movement style for conveying behavior character: for example, making scripted animation actions such as "pick up a glass" to be "neutral," "shy," or "angry" by altering the movement style [Amaya et al., 1996]. There are only a few such robot-specific projects to date [Harris and Sharlin, 2011, Saerbeck and Bartneck, 2010], none of which are able to learn style from demonstration as in our SBD systems.

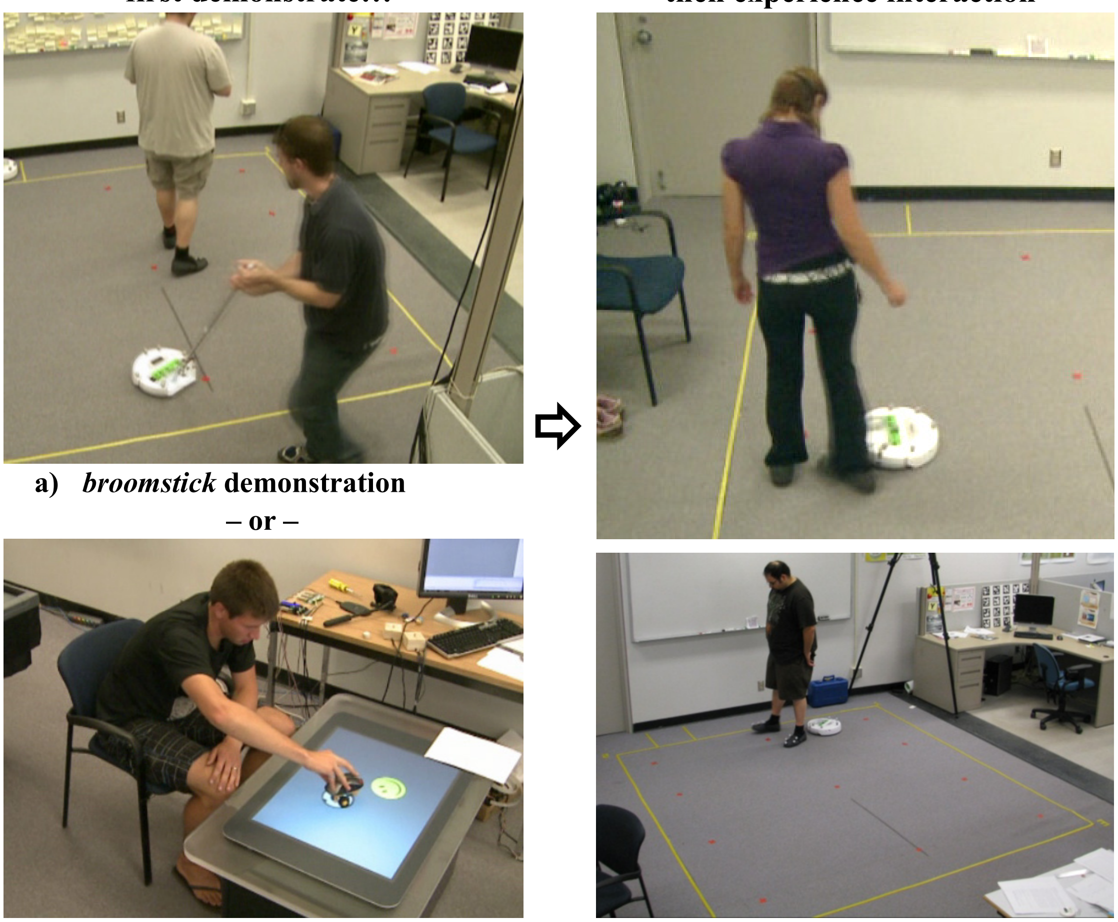
Programming by demonstration has been successfully employed since the early days of robotics [Halbert, 1984] for such applications as learning navigation routes [Kanda et al., 2007] or specific physical tasks [Gribovskaya and Billard, 2008]. Some robots such as Topobo [Raffle et al., 2004] or Curlybot [Frei et al., 2000] demonstrated the success and importance of enabling people to create goal-independent stylistic robot behaviors. These, unlike our Puppet Master SBD system, provide only a static replay of demonstration and the resulting behaviors are not interactive. Other programming by demonstration robot projects use style and emotion-charged elements as part of the demonstration-task interaction support: Breazeal et al.'s Leonardo robot uses facial expressions and style-laden gestures, while being taught, to convey such messages as lack of understanding or surprise [Breazeal et al., 2004]. Here the stylistic motions are not learned but serve as communication tools; the tasks being learnt are goal oriented. Our work extends the success of programming by demonstration by creating robots that can learn interactive locomotion style.

Due to the task and physical feasibility foci of most robotic programing by demonstration, related experimental evaluation has generally targeted technical goal-oriented measurements such as accuracy or task-completion success (e.g., [Breazeal, 2002, Matsui et al., 2005]) – these methods and results do not directly apply to the SBD research presented in this paper. In the broader domain, there is increasing evidence that human-centric and social aspects of HRI, such as is core to our SBD work, are particularly prominent in interaction and must considered in evaluation [Bartneck et al., 2007, Short et al., 2010, Young et al., 2009, Young et al., 2010c]. The quantitative method alone of distilling these complex human-oriented aspects into a set of statistical numbers is insufficient for properly describing the interaction [Strauss and Corbin, 1998], and so qualitative and exploratory evaluation methods often serve as the primary element of otherwise controlled studies (e.g., [Forlizzi and DiSalvo, 2006, Sung et al., 2009, Sung et al., 2007]), where the methods are used to describe interaction and to construct grounded interaction theories. Although less common for programming by demonstration specifically, existing qualitative evaluations explore and describe the interaction experience itself, as a way of building understanding of how programming by demonstration can be employed, for example, integrated into educational tasks [Frei et al., 2000, Raffle et al., 2004]. In our work we follow this precedent of using exploratory, interaction-experience oriented methods, and present the first such evaluation of SBD.

3. SBD INTERFACES FOR STYLISTIC ROBOT LOCOMOTION

Below we present our interfaces that we designed to enable people to author and personalize robot behaviors, by providing a demonstration of the desired interaction style. To realize robotic SBD, one requires interfaces for demonstrating exemplar interactive style to the robot, and a robot that can use the learnt stylistic interaction and an environment where people can interact with it (Figure 1). That is, first, a person uses an interface to provide a demonstration to the robot of how they would like it to interact, and

first demonstrate… **then experience interaction**



a) *broomstick* demonstration

– or –

b) *surface* demonstration

c) – interactive test bed

Figure 1 – Our SBD workflow, *broomstick* and *Surface* demonstration interfaces, and interactive-environment stylistic locomotion test bed.

following, the robot must be physically able to interact with the person to reproduce the style, e.g., to approach or follow appropriately, in real time. The actual procedure used in our SBD systems is straightforward: 1) a person provides a *single* demonstration of their desired behavior using one of the interfaces designed below, and is asked to finish as soon as they feel they have demonstrated all aspects of the behavior (no minimum time frame or repetitions are required), 2) output robot behavior generation happens immediately without human-involved or lengthy pre-processing.

As our current SBD incarnation targets the style of a robot's locomotion movement only, the interfaces detailed below only address this element of a robot's behavior. First, we detail our interactive-environment test bed where people can interact with and test the robots' stylistic locomotion in a reasonably large space (Figure 1c). We further detail two demonstration interfaces for acting out interactive locomotion styles: the *broomstick* demonstration interface, where a robot-on-a-broomstick is used to show interactive locomotion style (Figure 1a), and the *Surface* demonstration interface, where a tangible puck on a tabletop Microsoft Surface computer is used (Figure 1c). Both demonstration interfaces are designed to physically constrain the person's input to movements that are reproducible by the robot: the interfaces, like the robot, can turn on the spot but cannot move sideways. These inherent physical constraints "force" the person to express their desired movement style using the robot's actual movement capabilities and limitations.

One key but perhaps nebulous point regarding our SBD systems is that demonstration requires paired movements to enable the robot to capture the interactive component of the behavior: an example of how a person may move, and an example of how the robot should interact with the given person's movements. If we do not include the person's movement the robot does not have a reference point from which to learn the *interactive* component of the demonstration: it would only be able to reproduce the given static path, and not the appropriate real-time reaction to a person's interaction and movements. Thus, our demonstration interfaces include methods to provide both an example person movement and example robot interaction. This interactive behavior is similar to a *stimulus-response* system proposed in prior related work where a user specifies which response should happen given a specific stimulus [Wolber, 1997]. With stimulus-response the user consciously chooses and provides an example of a stimulus, and demonstrates an appropriate response to that stimulus. While fundamentally this is what is happening in our SBD system, we abstain from using this terminology directly for our work as we believe that there is a conceptual difference from the perspective of the user: in our SBD, the user simply provides an acted out continuous higher-level example of the behavior without necessarily thinking on the stimulus-response level.

3.1. The Characteristic Robot Sounds Extension

In addition to our primary goal of SBD for interactive robot locomotion, the SBD robotic Puppet Master algorithm used further allows users to demonstrate *when* robots should produce characteristic sounds [Young et al., 2010a]: for example, how and when to utilize and produce a *happy* beep sequence for welcoming an owner home. Here the style element is in knowing *when* to play these sounds as part of dynamic interaction. Sound was selected in this case for two specific reasons: a) sound is easily producible by our robots in comparison to, for example, facial expressions, and b) pre-programmed sound sequences are discrete actions (in contrast to the dynamic locomotion path) that serve as a proof of concept for non-locomotion-path extensions. If the robotic Puppet Master algorithm can learn how to dynamically use pre-programmed sound sequences as part of its interactive locomotion behavior, then any discrete action could be added in the same way: for example, when to take a picture, when to execute a pre-programmed "pick up object" command, or when to generate a happy face. Although this paper's main focus is on the robot's interactive locomotion path, we integrate sound – and thus the addition of arbitrary discrete actions – into both our interfaces and evaluations. For this work we selected a *happy* and a *sad* sound as initial proof-of-concept examples.

3.2. Interaction-Environment Test Bed

Rather than having a person interact with the stylistic locomotion results using on-screen simulations or by observing the interactive movement remotely, a primary goal of our interaction-environment test bed was to enable a person to directly interact with the robot; using a real robot instead of a virtual simulation can have important implications on interaction (e.g., as in [Guo et al., 2009]). We utilized an open space (Figure 1c) where the person interacts directly with the robot in real time, and where the robot interactively conveys its specific locomotion style towards the person. Using this test bed, a person can act out their role (e.g., a burglar entering a home) to evaluate the interactive robotic behavior (e.g., it acting aggressively toward the burglar).

Both the robot and person are tracked in real time using a Vicon camera motion-capture system. The robot is tracked via markers placed on top of it, the person wears specially-marked shoes (Figure 1c), and the control software directs the robot remotely using a Bluetooth connection [Young et al., 2010a]. The robot is an iRobot Create, essentially a simplified Roomba without a vacuum mechanism; the Create can turn on the spot, move forward and backward, but cannot move sideways.

3.3. Demonstrating With the Broomstick Interface

Our broomstick demonstration interface is a standard broomstick attached to a robot (iRobot Create, Figure 1a) that enables the demonstrator to directly show their desired interactive behavior style to the robot [Young et al., 2010a]. While one person walks around the space the demonstrator uses the broomstick to manipulate the robot to interact with that person in the desired fashion. The combination of the person's example movement and the demonstrator's example robot reaction constitute the demonstration data for the Puppet Master learning algorithm, and the resulting learnt behavior can be interacted with and evaluated using the test bed described above.

The design motivation behind the broomstick interface was to enable the demonstrator to express their desired interaction style as freely and directly as possible. Grasping the robot itself for control is not feasible due to the robots' small size and being low to the ground. The person cannot act their desired interaction behavior using their own body as this would incorporate many degrees of freedom not reproducible by the robot. The broomstick solution provides a fairly-direct demonstration method, while the robot on the end constrains demonstrator input to motions reproducible by that robot. The broomstick is rigidly bolted to the robot using a two-axis swivel, such that the stick itself can be freely tilted left-right and front-back to adjust the pushing and pulling vantage point, whereas rotating the broomstick directly rotates the robot: rotation on the spot is possible but cannot be done quickly due to the small radius of the broomstick and the resistance of the robot's wheels, although and slight rotation force while pushing or pulling intuitively steers the robot. Movements are tracked using the Vicon camera motion-tracking system, and the robot's wheel gears were removed reduce the force required to move the robot.

We attached soft-press buttons on the broomstick (Figure 2), that the demonstrator can use to indicate when the robot sounds (*happy* or *sad*) should be triggered during interaction: when pressed during demonstration the attached robot makes the appropriate noise as feedback. The buttons communicate with the host PC using a modified Phidgets wireless clicker kit.

Figure 2 – Broomstick buttons for triggering *happy* or *sad* sounds.

3.4. Demonstration With the Surface Interface

The Surface demonstration interface (Figure 3) uses a digital tabletop (Microsoft Surface) for demonstration. The example person movements for the robot to react to – this is an actual person for the broomstick case – are provided as pre-scripted sequences displayed on the tabletop as an animated happy-face, and the demonstrator uses a tangible puck (Figure 3) to show the desired robotic interactive movement style in relation to the animated person.

We designed the Surface interface explicitly for the studies presented in this paper, to explore a parallel demonstration approach compared against the broomstick interface: while the broomstick interface involves two collocated people (demonstrator and person the robot interacts with) interacting collaboratively using the actual target robot in the real-life target space, the Surface interface aims to simplify this into a scaled-down and perhaps more relaxed, seated scenario. The design of the Surface interface was an explicit attempt to address concerns that emerged from previous tabletop SBD work [Young et al., 2008]: we used the relatively small Surface such that the demonstrator can easily reach over the entire space while maintaining smooth motions, we provided the scripted person movements – rather than having two pucks – so the demonstrator can focus on creating one movement at a time without distraction, and the puck is tracked from the bottom to avoid problems with the demonstrator occluding top-mounted markers, e.g., as with the Vicon system [Young et al., 2008].

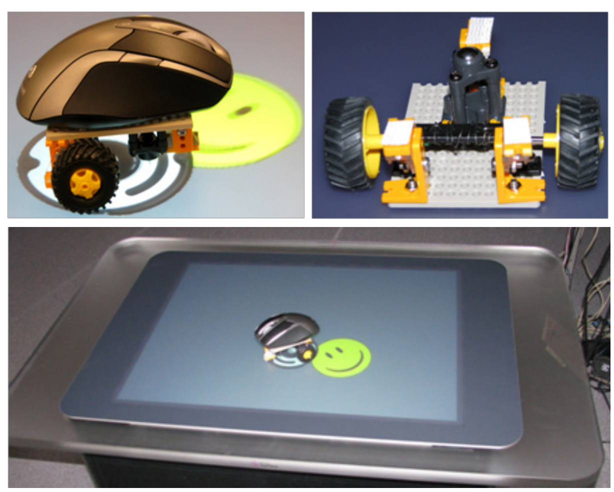


Figure 3 – Hand-held tabletop puck, top and bottom view, on Microsoft Surface. Reflective markers on bottom for tracking.

We designed the puck using a separate-axis wheels and caster design (Figure 3) to match the movement constraints of the target iRobot Create robot: it can turn on the spot but is unable to move sideways. Slight tension was added on the wheels using wire to restrict rapid movements not reproducible by the actual robot. The puck is tracked using reflective markers attached to the bottom (Figure 3), detected by the Surface's cameras as touches, and a wireless mouse was mounted on top to provide comfortable grabbing. The mouse's buttons provide a means to trigger the robot's *happy* or *sad* sounds, with audio produced during demonstration by the Surface computer.

4. THE ROBOTIC PUPPET MASTER ALGORITHM

Here we outline the underlying robotic Puppet Master algorithm behind our SBD learning and real-time behavior generation; full details can be found in the original papers [Young et al., 2008, Young et al., 2010b]. We present the algorithm here to build a general sense of the workings, capabilities and limitations inherent in our system.

Puppet Master's core workings revolve around a two stage process: 1) a rapidly iterating pattern matching algorithm (15hz) comparing the current (during output generation) situation to the demonstration data and 2) a frequency-analysis approach to behavior output generation that separates movements into general trajectory (low-frequency) and movement style detail (high-frequency), as a way to maintain style despite robots' rigid constraints.

The pattern matching component converts the data (both the demonstration and on-going real-time generation data) into a set of features based on relationships between the robot and the person, for example, distance between the two, relative position (behind / in front, or to the left / right) and relative angle (facing each other, facing away, etc.). These features are then considered over the time dimension (a 1s history window) to capture the derivatives, including one entity approaching another, turning away, or not moving for some time. The entire training data is compared to the current person-robot situation to find the best match instance – the target state – which is then used to generate the next robot output. Thus the robot behavior generation is a kind of patch work of the demonstration data (Figure 4). No underlying robot behavior or following system exists, the robot's entire movement model is extracted from the training data: without training, the robot does not move.

The comparison metric used is Euclidean distance, where the features above form scalar dimensions of a feature vector. The best-match training data is selected by minimizing the Euclidean distance between the real-time window of data and the moving window over the training data.

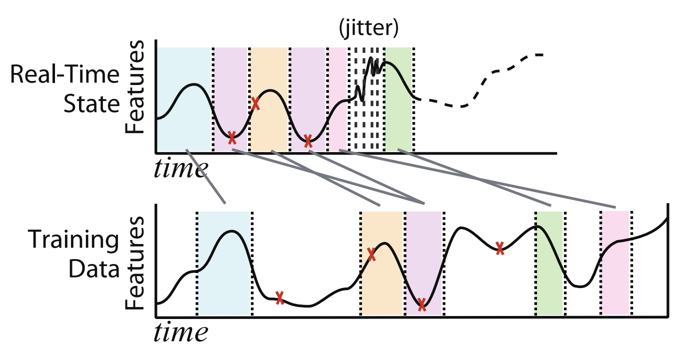


Figure 4 – Real time robot movement is a set of patches from the best-match training data (illustration only). All dimensions of training and output for both the human and the robot characters are summarized into a single curve for clarity. The red "x"s denote sound action events.

When to perform the *happy* or *sad* sounds was learned by associating the discrete actions directly with the training data on the time axis. This is denoted by the red "x"s in (Figure 4). Thus the sound actions are taken into consideration as part of the training-data similarity metric, and, when a patch of training data is used to generate output the associated sound is generated at that specific time.

The process happens at 15hz, where real-time human and robot movements are constantly being updated. This rate was selected through experimentation as being fast enough for interactivity (the robot can quickly respond to changes in human movements), and slow enough to reflect the practical movement limitations of the real robot: it cannot change its acceleration, direction, or speed quicker than this.

Exact reproduction of the target robot output movement (the patches) is generally impossible because of the feature set used: for example, the target may include relative human-robot positions far from the current real-time one, or the required movements to match target velocity may contradict the relative position. The robot itself must also adhere to rigid physical movement constraints such as limited acceleration or movement speeds. Puppet Master's solution is to apply rudimentary frequency analysis to the output path, to achieve a balance between proper position relative to the person, and the movement details – or texture. The robot is first steered toward the target's physical location, so that over time it maintains a fairly-appropriate relative position. Second, the relevant movement details from the training data are applied to the robot's path, such that while the robot tries to "catch up" to achieve its target location, it moves in the appropriate textured way, e.g., in a wavy or jerky fashion.

The general essence of Puppet Master makes it well suited to learning applications where few assumptions can be made about the behavior. By drawing directly from the training data it can conceivably reproduce any simple interactive movement style demonstrated. However, a limitation is Puppet Master's near-sightedness: it only searches over a small history during generation run-time, and there is no macro-level behavior model created to capture behavior trajectories over time. In addition, noise in the Puppet Master algorithm is manifested as occasional robot "jitter," where it will move rapidly toward quickly changing targets (highlighted in Figure 4). Despite these limitations, overall Puppet Master provides a robust and flexible SBD algorithm that has enabled the work presented in this paper, and serves as a first step toward more advanced SBD work.

5. STUDY APPROACH AND DESIGN

Above we detailed our SBD concept, the application to robotic locomotion, and our various interface designs for enabling the SBD workflow. In the remainder of this paper we investigate the feasibility of our approach by asking the following research questions: Q1) how do people respond to the core SBD concept of teaching robots to communicate using interactive locomotion style? That is, does it make sense to people to demonstrate movement style to robots? (user experience), Q2) Are our interfaces successful in enabling demonstration and evaluation of such behaviors? How do our interfaces differ? (interface usability), and Q3) how successful is our robotic Puppet Master learning algorithm in generating results that resemble the demonstration? (implementation efficacy).

In designing our study one problem we faced was the lack of similar systems against which we could compare ours. While there are many PBD implementations available, they either a) do not create *reactive* behaviors (results are static), b) do not target locomotion, c) do not focus on style but rather on task-level goals, or d) require detailed expert pre-processing and very large training databases. Most we examined faced several of the above issues inhibiting comparison. Therefore, rather than focusing on comparing our SBD interfaces and implementation against other systems, we rather qualitatively explore aspects of the users' experiences to develop insight into how people use SBD, and setup comparisons and independent variables within our system. This lack of prior work further created challenges in developing specific evaluation goals, and is a primary reason behind the exploratory nature of our work: we created broad study scenarios with simple goals (above) that would enable us to develop a better insight into the interaction and to form targeted questions for future study.

We conducted two interrelated studies (34 participants total) on our SBD interfaces: a demonstration study, where participants demonstrated behaviors to a robot and evaluated the results, and an observer study, where participants did not *author* but only observed and interacted with stylistic robotic locomotion. We further briefly present a programmer design critique that we conducted to compare the creation of stylistic robot locomotion by SBD to creation by more-traditional programming [Young et al., 2010a].

Table 1 shows the breakdown of how these studies fit together. The initial programmer design critique served the purpose of an initial experts-only feasibility check to see if people, even those with technical abilities, could use our SBD system to create stylistic, interactive behaviors that they are happy with (Q1,Q2). In addition, the results of this study provide programmer-created robotic behaviors against which our SBD can be later compared.

For the demonstration study participants create a range of behaviors using both interfaces (Q1). The study is setup around a comparison of the two demonstration interfaces (Surface and broomstick), between participants, and four different behaviors, within participants. This is a means to explore how the interfaces impact the SBD result and experience (Q2), how well people accept SBD across differing behavior types (Q1),

study	purpose	comparison variables
programmer	initial proof of concept *programmed* behaviours for later comparison	**within**: programming vs SBD for behaviour creation
demonstration	end-user authoring experience compare interfaces efficacy of algorithm	**within**: a range of stylistic behaviours **between**: *Surface* vs *broomstick* for SBD behaviour creation
observer	non-teaching perspective on the robot behaviours comparison of behaviours based on creation method	**within**: robot behaviour type and creation method compare entire results against *demonstration* study to explore impact of teaching

Table 1 – A breakdown of the three studies discussed in this paper, their purposes, and their comparison setup.

and how well Puppet Master learns various behavior types (Q3). Finally, we ask participants to identify their behaviors as a means to investigate learning quality, that is, if it captured the qualities of the behavior that the creator could identify (Q3).

The observer study removes the element of authoring completely and focuses rather on how the robotic SBD behaviors are perceived. The primary purpose of this study is to enable us to explore how the act of teaching impacts how people perceive the robotic behavior results (Q1). That is, this serves as a comparison point against the results of the demonstration study. A secondary point of this study is to enable us to investigate how the behaviors are perceived in relation to demonstrated intent (Q3), and also to compare how behaviors are perceived based on how they were created, as this study uses behaviors created by programming from the programmer design critique, as well as those created by each the Surface and broomstick interfaces (Q3).

In all of the studies presented here we selected four HRI scenarios with appropriate robot interaction styles for the participants to create: a robot politely following a person, a robot stalking a person, a robot that is happy to see a person, and a robot that is attacking a burglar. We refer to these shorthand as *polite*, *stalker, happy,* and *burglar*; participants were not given any further description and were free to interpret and demonstrate the behavior their own way. This set of scenarios was selected to roughly match those used in our previous animation-only effort [Young et al., 2008] as a range of behavioral styles. Although this selection was not grounded in behavior theory, and we admit that this limits our ability to generalize our conclusions of Puppet Master's specific learning abilities, we posit that the selection serves our primary exploratory purpose of creating a range of scenarios that engage the participant. It further provides a broad test of the robotic Puppet Master SBD capabilities: we remind the reader that the Puppet Master algorithm has no hard-coded behavior model, and must learn each of these behaviors completely and wholly from a participant's demonstrations. The specific count of four behaviors was chosen simply as a number that enables us to have some breadth while keeping the studies to a reasonable length.

To reflect our research questions, the guiding hypotheses of these studies are: we expect that people will naturally understand the "acting" concept of SBD for robots, will be able to use our interfaces to easily demonstrate interactive, stylistic behaviors to robots, and that the underlying Puppet Master algorithm will be technically effective in producing reasonable and satisfactory mimicry results that people can recognize and understand. We further expect that our results will reflect fundamentally on the use of SBD in general, above and beyond our specific instances and robot-locomotion modality. We hope to provide insight into how future HRI interaction techniques can be designed to effectively engage people with SBD; for example, how the interface should be designed, or what some of the trade-off implications of some related design decisions may be.

5.1. Programmer Design Critique

We previously conducted a programmer design critique to explore the preliminary question of how SBD differs from more-traditional programming techniques [Young et al., 2010a]. Four experienced programmers were recruited to first create interactive robot behaviors using a simple robot-control Java API and simulator, and then to create the same behaviors using the broomstick demonstration interface. Programmers were given 2 hours for the programming task, and no time limit for the SBD task: we limited programming to 2 hours to keep the study within a reasonable time frame, and from experience expected the SBD to take considerably less time. The driving questions behind this design critique included: would experts with programming knowledge still benefit from SBD? Where does SBD apply with skilled users, and, what are the trade-offs in comparison to direct programming? The participants were asked to create the four interactive locomotion-oriented behaviors: a polite follow, a robot stalking a person, a robot that is happy to see the person, and a robot that is attacking a burglar.

A primary result was that the programmers embraced the SBD approach: they anthropomorphized the robot and "played along" with the style-oriented character behaviors, despite their technical insight into the robot capabilities. Most critique results, rather than categorizing one method as being better, reveal differences between the two. Straightforward results were that all participants took the full two hours to program the four behaviors, and took an average of 10 m 27 s total in the SBD task, including iterative demonstration, observation, and simple discussion; exact times were 14 m 49 s, 8 m 52 s, and 7 m 40 s. Although one participant is excluded from the data as they requested not to be videotaped, we noted that they were not significantly different than the others. The SBD approach obviously saved time, a benefit that would be dramatically stronger for non-programmer users.

All participants further described a complex accuracy and control versus time and ease trade-off: programming gives explicit control but demands considerable time, and it is difficult to convert stylistic ideas into raw movement commands. SBD gives a direct, quick, and easy method for showing the demonstrator's intention at the cost of losing detailed control of the robot's movements. These trade-offs are not clear cut, as participants pointed out that programming gives a false sense of control, as: "even when I program I don't know exactly what is going to happen," and "when I see problems, I still don't know why it happens." Further, it was pointed out that SBD cannot be a perfect solution because the algorithm is "relying on its interpretations of [their] intentions, rather than on [their] actual intentions. There is no way to directly convey intentions."

Overall, this study suggests that the SBD approach may be useful even for users who have the capability to program behaviors more explicitly; it still makes sense to leverage people's demonstration abilities. Further, it highlights important limitations and benefits in comparison with traditional methods, insight that can be useful for both developing future SBD systems and for evaluating how people interact with them.

5.2. Qualitative Evaluation Approach

Given that there is not an established body of SBD work for stylistic locomotion that we can compare our work to, we believe that at this point it is less meaningful to focus on quantitative measures such as learning accuracy or completion-time efficiency, and instead take an exploratory approach of investigating users' experience with such SBD in general. That is, rather than focusing on our particular SBD realization only, we aim to develop insight into users' interaction dynamics with these systems that can help drive and direct future work in this area.

Therefore, for data analysis and investigation of our research questions we rely heavily on qualitative evaluation methods to investigate our hypotheses. We specifically use participant self-reflection via, e.g., participants' verbal comments, long-answer questions and opinion-oriented Likert-like scales as primary data for analysis.

Our quantitative analysis relies on standard difference-of-means statistics (ANOVA, t-tests, etc) to investigate such things as behavior identification accuracy. We also use non-parametric statistics to investigate effects of our variables on the user-experience oriented Likert-like scale data – this is ordinal data and may not follow a normal distribution and so non-parametric rank methods must be used [Field, 2009].

Our specific qualitative method draws heavily from the teachings of grounded theory: we performed open coding on the written answers, interviews, and video data, we analyzed the codes on varying levels to identify emergent themes and relationships throughout the data, and we used exemplary text quotations to clearly represent the complex ideas [Bernard, 2000, Strauss and Corbin, 1998]. The codes themselves have been removed from the data, but our presentation method uses short thematic titles to represent the emergent themes. Finally, we did not perform inter-coder reliability tests as there was only a single coder for the data.

We highlight that the validity of this approach lies in our exploratory research aims. Rather than attempting to make concrete and absolute claims about this initial-attempt immature technology, we rather use our technology as an opportunity to explore the broader interaction context and overall approach. Thus, we use the results as grounding for emergent themes and findings, culminating into a set of concrete SBD design implications for directing future work in this area. There are many recent examples of similar approaches in the community (e.g., [Voida et al., 2011, Wahlström et al., 2011]).

5.3. Current Studies: Overall Study Design

In all of our studies an experimenter performed the role of *actor*, providing fixed example movement paths for the robot to interact with (Figure 5). For demonstration, this enabled the participant to focus only on how the robot should interact with the actor, and for observing the participant would simply watch the robot interacting with the actor. This arrangement further ensured that the actor's movements were consistent across participants. We varied the acted paths (fixed across participants) at different study phases to avoid a sense of repetition, and in particular we used different paths for participant demonstration and result observation to avoid the perception of the robot simply replaying the demonstration verbatim. We carefully designed these paths to incorporate variety: short and long segments, turns, and stops.

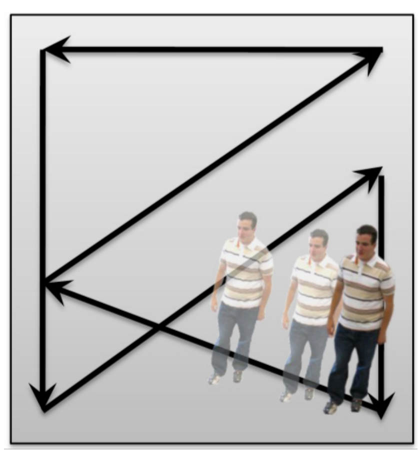


Figure 5 – Example scripted and acted movement path within the interaction space.

In this paper we are lacking concrete methods to measure similarity of behavior demonstrations, or quality of a generated result. Previous work with similar goals has developed related metrics for, e.g., comparing how "guessable" or "similar" pen-based gestures are to a trained target [Long et al., 2001, Wobbrock et al., 2005]. Unfortunately, for our work the problem is much more difficult than these previous examples where a static demonstration or result such as a pen gesture can be compared to an ideal gesture. The complexity stems from the paired human-robot demonstration and that the order of the demonstration has little bearing on the result. For example, two people may create a very similar behavior but go about it in very different ways: the order of their demonstrated features may be different, and one person 1) may provide a long demonstration where they repeat one important aspect several times and then give a single example of other key behavior aspects, while another person 2) may simply give a quick but clear demonstration with one example of each feature.

While this limitation of our work does hinder aspects of generalizability of results and comparability of the demonstrations and results themselves, we maintain that this is not necessary for our short term research goals of exploring user experience. This remains, however, important future work.

6. DEMONSTRATION STUDY

The primary goals of the demonstration study were to test if people accept and understand the concept of teaching style to robots, if our SBD interfaces are usable, and if the Puppet Master algorithm behavior generation results are satisfactory. As the studies in this paper are the first to target robotic SBD, in addition to the core questions above we take a heavy exploratory approach to both the study design and analysis, using open-ended questions and video data analysis.

The study design is as follows. Participants used either the broomstick or the Surface interface to author a set of interactive, style-oriented robotic locomotion behaviors, and we manipulated two independent variables: the interface used and the behavior created. The interface used, Surface or broomstick, was manipulated evenly between subjects (randomly assigned) and thus each participant only used one interface type. The behaviors that the participant created, *polite*, *burglar*, *happy*, and *stalker*, were fixed in order and manipulated within subjects: each participant performed the demonstration task four times, once per behavior type. These two independent variables (interface type and behavior) were chosen to directly consider the differences between the demonstration interfaces, and how participant reactions to SBD may differ for different robot behaviors. We note that having the behaviors fixed in order across participants introduces a potential confound of learning between the behaviors, and thus limits how much we can conclude regarding the differences of behaviors; however, this does not diminish our ability to analyze the overall use of SBD, or the differences between the interface types.

For the demonstration study we recruited 22 participants from the general university population, and each participant was paid $20 CAD for the one-hour study; they were 11 female / 11 male and aged 19–34 (M=26.9).

6.1. Tasks and Procedure

The demonstration study had two phases: behavior demonstration and identification of created behaviors. In the demonstration phase the participant created all four behaviors by SBD, and in the identification phase they observed their behaviors and attempted to identify them. For all components the actor's movements were pre-scripted (for reasons detailed in Section 5.3), or the on-screen happy face for the Surface interface only.

Participants completed a pre-test demographics questionnaire that also asked about predisposition toward robots, artistic experience, and technical ability. Before starting we performed a generic non-style example of the robot learning how to follow the actor that illustrated the overall process, and allowed the participant to physically handle the assigned interface (broomstick or Surface) to ensure they understood how it worked.

The behavior demonstration phase was conducted as follows (Figure 6): the participant first demonstrated a given behavior (e.g., *burglar*) using their assigned interface (i.e., Surface or broomstick); they manipulated their interface in real time to the actor's / pre-scripted happy face's movements to demonstrate their desired interaction style. The participant had full control over the robot as it had no prior behavior and learned the entire interaction style from the participant demonstration, and the participant was not told ahead of time how many behaviors they would create or what the specific types were. The participant then observed the resulting robot behavior interacting with the actor on the test bed, and was asked to reflect on if the behavioral style was in accordance with what they were attempting to create. Participants were allowed to observe for as long as they wish, where the actor simply repeated their motion path, and the participant indicated when they were done observing. If they decided that the results were unsatisfactory, the participant could attempt to re-demonstrate the behavior as many times as they wish. Given satisfactory results, they completed a questionnaire regarding demonstrating the given behavior before continuing on to the next behavior type. We recorded the training and observation times and number of training-observation cycles for analysis. Note that for the broomstick interface participants performed both demonstration and observation in the same space, where the Surface interface required participants to change spaces.

For the behavior-identification phase we presented the behaviors just created in a shuffled order (fixed across conditions and participants) using the test bed and the actor,

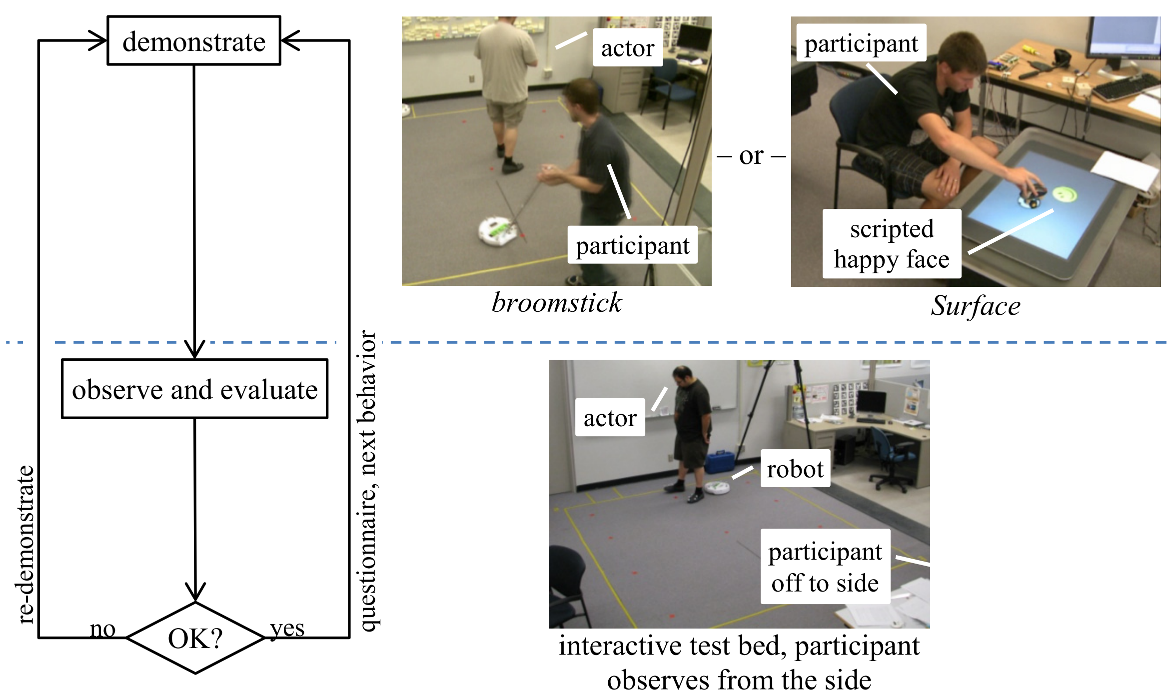


Figure 6 – The behavior-demonstration phase workflow.

and participants attempted to identify them. Observation time was not limited but the participant could not return to a previous behavior once they moved forward. The post-test questionnaire enquired about the overall experience using a mixture of Likert-like scales and open-ended reflection questions.

6.2. Quantitative Results

Demographics – No correlation was found between artistic ability, or experience with programing or robots, on training time, identification success, or any other measure tested.

Training and Observation Time – The grand mean of training time per behavior across all cases was 50s (SD=37s, min=4s, max=261s), and observation time was 115s (SD=68s, min=24s, max=450s). Participants took on average 1.27 tries to train their behavior (max=3), with 9 of the 22 participants never re-training. Mixed-design ANOVAs (with logarithmic time transform to improve normality) found no effects of interface (broomstick or Surface) or behavior type on training or observation time.

Per-Behavior Questionnaire – Figure 7 is a frequency table of the responses to per-behavior questionnaires given after behavior training. No effect of behavior type was found (Friedman's ANOVA) on the first and last questions. There was an effect of behavior type on how "mechanical" ($X^2(3)=16.43$, $p<0.001$) or "natural" results were ranked ($X^2(3)=9.51$, $p=.023$). Post-hoc pair-wise Wilcoxon Signed Ranks tests (Bonferroni correction, 6 pairs, significance at $p=0.008$) showed that participants ranked *polite* as more "mechanical" than *burglar* ($Z=-3.09$, $p=.002$, $r=-.68$) and *happy* ($Z=-3.21$, $p=0.001$, $r=-.69$); no other effects were found. Although post-hoc tests found no significant results for "natural," Figure 8 shadows the results for "mechanical."

Behavior Matching – Figure 8 shows behavior identification results: average matching success rate 67%, SD=37%, max=100%, min=0%. Half of the participants identified 100% accurately. No effect of behavior (Friedman's ANOVA) or interface (Mann-Whitney Test) was found on identification.

Post-test Questionnaire – Figure 10 presents a summary of post-test questionnaire results. Mann-Whitney tests did not reveal an effect of interface type on any question.

6.3. Qualitative Results

Here we present themes that emerged from the open-coding and exploratory analysis of participant feedback, video data, and observations.

Robot-Human Collisions – Concern over robot-actor collisions was very prominent, both in the verbal and written feedback. Several stated "the robot needs some underlying assumptions like `don't drive over person' [that] should take priority over demonstrations," and some suggested safer ways to be aggressive, for example, "maybe aggressive motions would be too much of a threat/danger, but the angry sounds would be good."

Movement Jitter – The most frequent complaint was that the robot's movements "seemed really jerky" and were often "too abrupt" – this was an artifact of the learning algorithm. While some used technical language such as "the robot made too many turns back and forth," most participants used behavior-oriented or emotive language to explain it, e.g., that the "robot seemed a bit confused" (the most-common attribution by far) or that "the robot looked like [sic] thinking and deciding." Many related this observation

	strongly disagree 1	disagree 2	neither agree nor disagree 3	agree 4	strongly agree 5
you were satisfied with how well the system captured the behavior you were trying to demonstrate	0	8	10	54	16
the resulting robot behavior felt overly mechanical	4	17	36	27	4
the resulting robot behavior felt natural, organic, possibly human controlled	2	20	31	26	9
I think it makes sense to teach robots this behavior by demonstration	0	6	15	43	24

Figure 7 – A table of per-behavior post-training question answers (4 behaviors across 22 participants).

	strongly disagree 1	disagree 2	neither agree nor disagree 3	agree 4	strongly agree 5
polite follow	2	7	6	5	2
stalker	0	7	7	7	1
burglar	0	4	7	8	3
happy to see you	0	2	11	6	3

Figure 8 – Participant responses to "the resulting robot behavior felt natural, organic, possibly human-controlled".

	polite follower	stalker	attacking a burglar	happy to see you
polite follower	14	5	1	2
stalker	6	14	0	2
attacking a burglar	1	0	17	4
happy to see you	1	3	4	14

Figure 9 – Frequency table of participants matching their own behaviors, the diagonal is the correct match.

	strongly disagree 1	disagree 2	somewhat disagree 3	no opinion 4	somewhat agree 5	agree 6	strongly agree 7
I enjoyed the overall demonstration and observation of behaviors	0	0	0	0	0	10	12
I was often disappointed by the resulting robot behaviors	2	6	6	2	5	1	0
The process of demonstrating to a robot was often frustrating	4	12	5	1	0	0	0
The resulting robot behaviors conveyed the style and personality I intended	0	0	0	0	10	11	1

Figure 10 – Frequency table of post-test questionnaire answers.

back to a lack of learning ability: perhaps the robot "did not fully understand what to do," and one participant suggested that the robot "might need more sensors to stop seeming confused." One person stated: "if I didn't know the behaviors the robot is mimicking, I would say it's trying to act confused."

Taking Personal Responsibility – Participants commonly took responsibility for quality problems: "maybe it was my fault as a demonstrator," or "[I'm] not very satisfied. I think the reason was not very efficient demonstration." This also applied to robot jitter: "it was jerky, but then again, I was moving quickly during training," and related to responsibility: "what if the robot doesn't learn? Will it hit/harm the human/itself?"

Reactions to the Broomstick Interface – Reaction was overall positive, for example, "it was handy and helpful in movements," and several participants claimed it was a "good idea." One requested "a body suit for the performer" that would track the person directly, but conceded that "regarding the type of student robot, [this] was the best way of control." Feedback also pointed to areas for improvement: some participants said "it was a little bit hard to demonstrate using a broomstick," for example, "it would be hard for me to make the robot rotate on the spot." Two participants raised concerns of how the broomstick's properties influence training, and thus the resulting behaviors, as "the broomstick causes the movement to be a certain way," adding "inconsistency to the robot's behavior, making it indistinguishable whether the robot has learnt the poorly-performed behavior or did not learn appropriately." One person noted confusion as both the robot (a vacuum model) and the broomstick have a strong image of cleaning.

Reflections on the Surface Interface – The primary complaint was regarding the physical size of the table: "the task space should be larger / puck be smaller." One participant (who trained using the surface interface only, and was unaware of the broomstick interface) asked if they could "maybe demonstrate to the robot visually? I feel that demonstrating or teaching the robot directly may be more effective." Further, several requested to "include barriers and items inside the space," (none exist in the demonstration area) particularly as "the stalker effect isn't noticeable without barriers and obstructions." This request for physical barriers was not mentioned for the broomstick.

Attitudes Toward Teaching Style to Robots – Participants primarily expressed approving sentiments such as they wish "more studies like this were conducted, because some of existing robots really lack in the human interface quality," that "robots need to be trained quickly and easily if we are to implement them in every home," and as "each person's interpretation of aggressive would be different, it wouldn't make sense to pre-program the behavior." One participant noted that "when [intended behaviors] are ambiguous (e.g., what is \`excited' anyway?), it's a good idea." No participants expressed negativity toward the idea of training a robot, although some worried that the given behaviors "feel too complex to learn in a short period of time," and some noted that "it was easy to identify distinct tasks but hard to identify similar tasks."

Many participants (15) reported post-test that they would care about the personality and style of the robot; there was no difference between those that used the Surface or broomstick. A few participants were "not sure \`social' robots will be very important," for example, "I don't see a point in teaching this behavior [*happy*], but if you must, this method works," and one participant was "not looking for a pal," and another said "I don't think I would worry about \`personality'." Some expressed uneasiness, for example, "it would be scary when [a robot is] too natural and gets a `soul'" and "I'm kind of scared that robots can control humans someday."

Positive Response and Engagement – Participant feedback given both orally and in written answers was generally positive and enthusiastic: "robot followed the demonstration pretty accurately," "I am amazed!", "the robot did a good job in this case, better than I tried to do." Many participants expressed pride at being able to create robot behaviors: "when I saw all of [the behaviors] at the end, I was quite happy," and one person exclaimed that their *happy* behavior "can [sic] welcome the visitors while entering home." A few reported that they felt the robot behaviors improved as the experiment progressed, although this was actually not founded in the technology. Finally, we noticed that even people who openly expressed negativity toward the idea of emotion and robots used anthropomorphic language when discussing the robot.

Limitations – Some participants felt restricted by the robot's limitations and our locomotion focus: for example, "a dog would run around, jump, move its tail and follow its owner," or "my definition of `excited' resulted in high velocities which the robot was unable to reproduce." Others noted that the robot "can only do what is shown to it, no creativity," for example, "the robot reproduced too many details and not just the general idea," or "at the beginning, I had a hard time with moving my robot, the robot reproduced that as well." One participant noted that this "requires the teacher to be a *good* teacher."

Training Robotic Sound – While several participants expressed regret that they "totally forgot about sound!" others claimed "the robot sounds were necessary to understand the behaviors," and one person stated that "it's the only way to tell the difference between happy or angry." Several related sound to the overall character: "it brought a human dimension to the experience." Some requested more sounds, for example, "for the burglar an angry sound would have been useful."

6.4. Discussion

The demonstration study results paint a broad picture of participant interaction with our SBD interfaces and provide insight into how SBD was received. Primarily, the results validate the technical and conceptual success of our system: people readily understand SBD, our interfaces are usable for demonstrating to robots, and the Puppet Master algorithm results were acceptable.

The results strongly show that people understand and accept the SBD approach: none of the participants was observed having, or reported having, problems understanding the underlying concepts or tasks, there were no complaints regarding the task of teaching, and many applauded the idea of easy customization. Many explicitly supported the subjective importance of robot interaction style. As we found no correlation between technical or artistic ability and any measure tested for, we believe that not only is the system accessible to non-experts, but that perhaps being better-trained in a relevant area may not impact the ability to use our SBD interfaces effectively. Perhaps the reason for this broad acceptance is that SBD leverages the widely-accessible *social stock of knowledge* [Berger and Luckmann, 1966].

Our results show that the robotic Puppet Master algorithm enables people to quickly (M=50s) create interactive, robot locomotion behavior styles that they are satisfied with and could reasonably identify: half of the participants perfectly identified their behaviors.

The overall 67% behavior identification rate was less than hoped for as we anticipated near-perfect given the broad spectrum of styles; this perhaps highlights the need to improve the Puppet Master algorithm to better capture and reproduce behavior elements. We note that a future point of analysis could be the similarity of a given behavior between participants, and the similarity of behaviors for a given participant. This could provide insight into how people modulate their demonstrations between behaviors, and, if there are global similarities across participants; this information could be useful for future improvement of the underlying learning algorithm. Further, we informally note that the experimenters were often surprised that participants could not identify behaviors that, to the experimenters, seemed quite obvious. We attribute this to the difference between more experienced observers and the first-time participants who may not have carefully considered what they were demonstrating or watching for.

Both the broomstick and Surface interfaces were successful in their goals of enabling people unhindered and easy demonstration of interactive robotic behaviors, while maintaining robot-specific movement constraints. However, the lack of difference in quantitative results was surprising. We believed, for example, that the Surface may be faster for training given its much smaller, more manageable interaction space. We also assumed that the broomstick would produce better (i.e., easily recognized) results as its manipulation requires full body movement that may be more physically engaging, encouraging demonstrators to give detailed and elaborate input. However, the behavior type appeared to be a stronger factor on our findings than the interface type. On the surface we see this as a reflection of the success and generalizability of both interfaces, but this raises the question of “how generic interfaces should be in terms of behavior types?” Would there be any benefit to creating more specific interfaces, for example, one non-generalizable interface to demonstrate *stalker*, and another for *burglar*?

Many participants did not use robot sounds despite expressed interest and claims of necessity: we are not sure of whether this is due to interface design problems (e.g., perhaps the sound usage was not obvious or easy to use in both interfaces), or that the idea of demonstrating sounds to a robot in this context may not make sense to people. As this emerged from the data analysis we were not able to inquire with the participants on further reasons for this. (We informally noted that Puppet Master was reasonably successful at integrating sounds into the generation). Further studies would have to be conducted to further explore this result.

Apart from primary findings discussed above, below we present themes of interest that emerged from the analysis.

People Understand Teaching – The data showed that people are perhaps more adept at casual teaching than expected, and clearly understand the underlying intricacies. We were impressed by the depth of understanding, for example, some acutely noticed the robot cannot directly distinguish between intention and actions, that it has “no creativity,” and that SBD requires a “good teacher” that understands the robot’s needs. Participants further understood that the properties of the training system (broomstick or Surface) impact the training and resulting style. Our overall successful results suggest that this level of understanding helps users to work within and around the inherent complexities and limitations of robot teaching to produce satisfactory results. We note, however, that this does not necessarily make the person a *good* teacher who knows which behavior characteristics are important or key to the overall *style* (a concept highlighted in prior work where participants were found to not necessarily select appropriate components or objects to define a behavior [McDaniel and Myers, 1999]).

Concern for Robot Actions – Concern was stronger than expected for the robot physically touching the actor during test bed observation, despite it being clear that no threat was posed. Particularly interesting is the *burglar* case where participants explicitly trained the robot to attack (and usually to collide with) the actor or smiley face icon; no difference was discovered between the Surface or broomstick cases. In addition to highlighting the issue of robot safety, this finding shows a difference between how people respond to the idea of a robot attacking a person (i.e., during demonstration) versus seeing their own aggressive robot behavior design attack a person during observation.

Attachment to Behaviors – The excitement and pride that people showed for their behaviors, as well as potential responsibility for the robot's actions, raises questions regarding the impact of enabling people to customize their robot. SBD researchers should consider such questions as: does teaching affect usage or perceived robot success? Or alleviate issues of fear, worry, or unease with the robot? This relates to the concept of the *extended self* in relation to possessions [Belk, 1988, Kiesler and Kiesler, 2004].

Jitter Perceived as Confusion – The degree to which people interpreted the jitter as a robot personality trait (such as being confused or changing its mind) was surprising: we find it serendipitous that underlying uncertainty in the Puppet Master algorithm, manifested as robot jitter, was accurately and naturally interpreted as confusion. Robots can thus use jitter to communicate uncertainty in anthropomorphic terms.

Anthropomorphism – As in prior HRI work, participants naturally tended toward anthropomorphism, in this case, without anthropomorphic design and even for the very short interaction span; for a few, this was despite negative predispositions to the concept of robots with emotions, e.g., the idea of "happy" robots. This parallels work that found that people treat media and machines as living things despite their often explicit and adamant opposition to the idea [Reeves and Nass, 1996].

Behavior Selection – Participant matching of behaviors (Figure 9) suggests a clustering of matches, such as an overlap between *polite* and *stalker*, and the *burglar* and *happy*. As this is similar to the results found in the animation-only study [Young et al., 2008], it is not clear if the clustering is a result of the underlying behavior choices, or the particulars of the Puppet Master system's applicability to behavior types. This highlights the importance of considering the testing behaviors carefully, and we recommend future-work exploration into existing theories on personality types and emotion.

The results of the demonstration study support our SBD approach, interfaces, and the underlying Puppet Master algorithm as an effective end-user system for the creation of interactive, style-oriented robotic locomotion behaviors, and further uncovered key findings and recommendations for future robotic SBD research and interface design.

7. OBSERVER STUDY

The demonstration study above focused on the SBD aspect of our work: how participants can use our overall system to create interactive style-oriented robot behaviors. The observer study, in contrast, only has participants observe and interact with the robot behaviors. The main purpose of this study was to provide a non-teaching comparison point against the demonstrator study, as a means of investigating how the act of teaching impacts interaction. In addition, we use this study to test the underlying Puppet Master behavior learning algorithm: the robot behaviors are taken from the demonstration study, and we see how observer reactions to the behaviors match the original intent. Finally, this provides an opportunity to explore people's reactions to and acceptance of stylistic robot locomotion behaviors outside the teaching context.

We recruited 12 participants from the general university population, aged 19-36 (M=26.3), 7 male and 5 female, whom we paid $20 CAD for one hour. Observers were videotaped with permission (half requested not to be), the footage used for analysis. Participants were encouraged to *think aloud* regarding their experiences and thoughts.

7.1. Tasks and Procedures

In this study participants observed a robot interacting with our actor (see Figure 6, right side), where the robot exhibited several behavior styles in turn. We used the 4 behavior types (*happy*, *polite*, *burglar*, and *stalker*) as an independent variable of the study. The other independent variable was the source of the behavior, 3 levels: two were from the demonstration study created by the 1) broomstick or the 2) Surface interface, or created by 3) programming methods (Section 5.1). The experimenters selected these behaviors subjectively as the best results from the prior studies. While we concede that without a concrete measure of quality this limits the analysis that can be performed on these behaviors, we maintain that this selection method is sufficient to help us investigate our primary research questions of user experience relating to SBD.

These behaviors were all shown to each participant (the independent variables were manipulated within subjects). Thus we explored if behavior perception is dependent on how it was created, in part as a means to test the effectiveness of the authoring methods. In particular, is there a significant difference in how observers respond to behaviors created by SBD or established-methods programming? For example, are programed behaviors more-easily identifiable?

The observer study had 3 phases (Table 2). First (exploratory phase) participants observed an instance of each of the 4 behaviors types (all created by SBD) for 4 min. while thinking aloud; participants were not yet aware of the behavior types. For the subsequent second task (identification phase) participants were informed of the behavior categories and asked to classify 12 additional behaviors after a 45s observation each: 4 each of behaviors created by broomstick, Surface, and programming. For all 16 cases the actor walked in the same pattern, the order of behavior presentation was counterbalanced, and the particular behaviors were subjectively chosen by the experimenters as being high-quality results from the demonstration and programming studies. The third task (in-situ phase) was unstructured, and the participant could choose to themselves interact with the robot exhibiting

phase	exploratory	identification	in-situ
behaviors	4 SBD	4 SBD *broomstick* 4 SBD *Surface* 4 programming	same 4 SBD as open-ended phase
goal	elicit overall reaction to robot behaviors	match observed behavior to intent	observe people interacting first-hand with robot behaviors

Table 2—The layout of the *observer study* three phases.

the four exploratory phase behaviors. This phase was optional, by verbal enquiry, as a means of measuring participant interest. This further provided an opportunity to observe participants interacting first-hand with the test bed interface.

The procedure was: we administered a pre-test questionnaire for demographics, predisposition toward robots, and artistic and technical experience. Participants observed all behaviors for the first task while performing the think-aloud, then continued to the second task where they observed each behavior for 45s and completed a quick questionnaire after each. After the optional third task we administered the post-test questionnaire.

7.2. Exploratory Phase Results

The exploratory phase elicited reactions from participants regarding the robotic behaviors, unbiased by opinions of what the behaviors should be: participants were not even informed that the robots have personalities or style-oriented behaviors. Participants generally used straightforward description regarding robot speed, angles, or proximity to person. Many cared if the robot "does a good job of staying in the boundaries." Anthropomorphism was common, for example, "it seems for me that when he's thinking about what to do, he beeps;" many people called the robot "he" (no one was observed using "she"). Only one person expressed animosity to the idea of robots having emotions or human-like personalities, although this person was quite strong. Below we outline participant responses to the particular individual behaviors:

Stalker Description – Some described using language such as the robot is "trying to hide, trying to follow," or "it is trying to avoid?", although participants remained largely silent, and several voiced confusion regarding what the robot was doing; there were very few comments for the *stalker* behavior.

Burglar Description – Observations included the robot is "trying to hit [the experimenter]," "trying to get ahold of him because it keeps jumping on him" or "it's aggressive, as if it's fighting for territory;" "aggressive" emerged as a theme. One participant mused that the robot is trying to say "`nothing here for you! What are you doing here?' maybe he wants to say `please pay attention to me'." There were no comments on the *sad* sound used in this *burglar* instance.

Polite Description – Participants attributed personality such as the robot "somehow [sic] look like a police man / security guard, walking around campus," or that it was "approaching the person more carefully" or that it "seems more polite this time." Some drew parallels to the *stalker* behavior, and other comments were simply descriptive: the robot was "trying to follow as closely as it can" or "does nothing but just follow him."

Happy Description – Some noted that the robot was "not quite as violent as the [*burglar*]," or that "it's in a good mood," "very opposite to [burglar]." Others found the robot "aggressive, but not as much as the [burglar]," or to "seem scared" or "nervous:" "after hitting [experimenter] seems to look frightened." Many people commented on the unclear meaning of the *happy* sounds, for example, it "sounds like when a battery is going low," or "to me, sounds very neutral, doesn't sound like a good sound or a bad sound." One participant said that they "found it a little hard to associate the sounds to certain behavior when [they] didn't know what the behaviors were."

7.3. Identification Phase Results

Following the exploratory phase, participants were informed of the behavior types and attempted to identify an additional 12 behaviors; results are given in Figure 11. Overall match success was 54% (SD=16%, min=25%, max=83%). A one-sample *t* test comparing our distribution to the expected mean of 25% given random noise (four choices) shows that it is unlikely that 54% was achieved by chance ($t(11)=6.13$, $p<.001$). No significant effect of behavior-creation method (*Surface*, *broomstick*, *programmer*) was found on match success. As with the demonstration study, clustering emerged with matching: Figure 11 shows how *stalker* and *polite* were often mistaken for each other.

Per-Behavior Questionnaire – Behavior type had an effect on responses to several per-behavior questions (Friedman's ANOVA, summary in Figure 12); no effect was found for "I felt like a human was controlling the robot" or "the behavior felt mechanical." Post-hoc tests were not practical as Bonferroni correction over 66 pairs would diminish the power of the test (significance at $p<.00075$). Rather, we present the average Friedman ANOVA's rankings (Figure 12) to provide insight into possible relationships for directing future investigation. We did not analyze the groupings of the 12 behaviors per behavior type or creation-interface type due to the difficulty of performing statistical tests on the two-way non-parametric dependent factors. [Field, 2009]

7.4. In-Situ Phase Results

Eleven of the 12 observer study participants opted to interact with the robot themselves for this open-ended in-situ phase. While the exception cited simply a lack of interest, the rest expressed excitement and were quite animated with interacting with the robot. Participants were by now aware of the behavior types, often requesting us to load particular ones, and readily played along, for example, by quickly moving away from the *burglar* robot with exaggerated motions when it was chasing them. Comments included "the way it moves, the sound, all makes it creepy like a stalker," (*stalker*) "the robot seems happy!", "`happy to see you' is just a pet of child who really feels happy to see me," (*happy*) and the robot "looks and feels like a polite machine." (*polite*) Further, several participants said during this phase that overall the robot "reminds [them] of a dog."

Regarding the specific participant who strongly voiced animosity toward the idea of robots having personalities (exploratory phase): once this person engaged the robot, they were laughing, talking to the robot as they may an animal, and now used anthropomorphic – in contrast to previously descriptive – language to describe the robot: "he's doing a good job" (*burglar*).

When asked for opinions on interacting with the robot, many participants re-iterated interest (e.g., "definitely wanted to"). Some used this time to test the behavior: "I found it very entertaining trying to predict the behavior of the robot and seeing how it reacted." Participants also reported that they "could get a better idea of some of the personalities when [they] interacted with it, compared to simply watching." Feedback on behavior quality was positive: "for the most part the behaviors seemed very natural and I was able to believe the robot had a personality of its own," or "the behaviors and displayed intelligence of the robot was very impressive."

A problem in this phase was that participants would move more quickly than the actor did, and the robot was too slow to interact properly with the participant: one complained

	stalker			burglar			polite			happy		
	tabletop	broom.	prog.	tabletop	broom.	prog.	tabletop	broom.	prog.	tabletop	broom.	prog.
stalker	9	6	10	2	3	0	1	6	6	2	3	0
burglar	0	1	0	4	1	10	1	0	0	2	2	1
polite	3	5	2	4	4	0	8	5	6	3	1	3
happy	0	0	0	2	4	2	2	1	0	5	6	8

a) grouped by behavior, correct responses on rows aligned with behavior type

	tabletop				broomstick				programmer			
	stalker	burglar	polite	happy	stalker	burglar	polite	happy	stalker	burglar	polite	happy
stalker	9	2	1	2	6	3	6	3	10	0	6	0
burglar	0	4	1	2	1	1	0	2	0	10	0	1
polite	3	4	8	3	5	4	5	1	2	0	6	3
happy	0	2	2	5	0	4	1	6	0	2	0	8

b) grouped by creation type, diagonals are correct responses

Figure 11 – How observers classified behaviors shown to them, 12 per participant.

	tabletop				broomstick				programmer				Friedman's ANOVA
	stalker	burglar	polite	happy	stalker	burgalr	polite	happy	stalker	burglar	polite	happy	
it was difficult to classify the behaviour	5.13	7.17	7.25	8.25	8.38	6.92	7.96	7.96	2.79	4.33	5.38	6.50	$\chi^2(11)=35.5$ $p<0.001$
I found the behavior to be engaging	7.67	5.88	5.04	5.79	3.88	7.42	4.92	5.33	8.29	9.71	6.83	7.25	$\chi^2(11)=32.8$ $p=0.001$
the behavior fell into the categories I was given	8.17	6.17	5.71	5.33	4.25	6.50	5.21	4.79	9.13	9.33	7.58	5.83	$\chi^2(11)=33.2$ $p<0.001$

Figure 12 – Average ranks of how participants rated each behavior on a given question, where higher number was stronger agreement; ranks are relative to each participant and indicate how behaviors were rated against each other.

they were "waiting for the robot to interact with [them]." Some felt that these behaviors "were not as natural" as some of the previous ones, although they were the same.

7.5. Post-Test Questionnaire and Overall Observer Study Results

Here we present themes that emerged from the open-ended post-test questionnaire.

Jitter as a Personality Trait – The predominant comment post-test and throughout the study was regarding robot movement jitter. While some of this was simply explanation, such as the robot "shakes a lot," participants usually anthropomorphized the robot jitter: it "seems to be very indecisive [sic] on the movements," "it seems frustrated when it jitters," it is "very ADD [Attention-Deficit Disorder], it gets distracted." Attributing "confusion" was particularly common, for example, "trying to follow but is very confused." One participant said the jitter is "somewhat a dog, he [robot] smells something some times."

Sounds – Several participants said "the sounds (beeps) were helpful" and "important" for identifying behaviors; one said: "when I was unable to determine the robot's behavior by its actions, I relied on the sound to determine if the robot was `happy' or `angry'." Others also noted that sound "is very important to give natural feeling to users." Several participants found that the "happy tone was not clearly happy," although there were no comments on the *sad* sound commonly used for *burglar*. Some participants reported that they did not pay attention to sound.

7.6. Discussion

The observer study served primarily as a non-teaching comparison tool against the SBD demonstration study, and clear differences emerged.

First, the issue of robot-human collisions, which was a strong theme in the demonstration study, did not emerge at all in the observer study although the robot was particularly aggressive in the *burglar* scenario. Further, only in the demonstration study were participants found making excuses for or trying to rationalize poor robot behavior; *observer* participants were much more vocal and negative regarding problems such as robot movement jitter. The act of teaching altered how concerned and critical participants were about the resulting behaviors, and thus perhaps teaching created a sense of personal responsibility for potential violence and mistakes. This may explain why demonstration study participants had increased tolerance for problems and rationalized them.

The observer study provides some support for the Puppet Master algorithm, that it captures sufficient personality traits for creating a better-than-random correlation between intended trained behavior and perceived result (54% versus 25% random). The simple fact that the SBD behaviors created in 50 seconds (average) by untrained members of the general university population can compete with behaviors programmed in roughly 30 minutes each by experienced programmers [Young et al., 2010a] speaks to the success of the SBD the approach and potential for the algorithm. However, there are still improvements to be made. For example, future work should formally explore the selection of behaviors, and how this may impact the algorithm design, for example, by applying previous work that mapped the dimensions of style [Gallaher, 1992] to Puppet Master as a way of systematically exploring and defining its strengths and limitations.

The results highlight that stylistic interactive locomotion can be used to create robot personalities and that people will readily attribute personality to such movement styles without other visual or morphological cues. In particular, observer participants had considerable "buy in" even without explanation such as received by demonstration participants, and without experience deciding how the behaviors should be conducted.

Below we discuss additional points and themes that emerged from the analysis:

Jitter Perceived as Confusion – Although participants in the demonstration study also used confusion to understand jitter, it emerged here without the context of teaching and learning. We believe this suggests an intrinsic anthropomorphic "confusion" meaning within the physical act of a robot "jittering."

Interacting With versus Observing – We feel the fact that 11 of the 12 participants expressed excitement and opted into the in-situ phase, even without tangible incentive, illustrates how engaging the robot behaviors were, even accounting for the novelty factor of robots. Our results further pointed to some differences between participants observing and themselves interacting with the robot, such as level of anthropomorphism and "buy in" with the behaviors. This suggests future work on, for example, considering how the meaning of behaviors changes for direct interaction versus observation.

Reception of Sounds – Unlike in the demonstration study, the robotic sounds used were not introduced to observers as *happy* and *sad,* and as such participants did not have predisposition toward their meaning. The *sad* sounds often used during the *burglar* behavior rarely received comment, so we make the assumption that they at least did not raise confusion, but overall participants expressed a lack of clarity regarding *happy* sound. This raises questions regarding the generic nature of sound sequences and the flexibility of interpretation within a context that should be considered for HRI design. Despite the ambiguity, many people still claimed the sounds were helpful for understanding a behavior, and added a "natural feeling" to the robot: this supports sound as playing an important role in shaping behavior meaning.

8. OVERALL REFLECTIONS AND IMPLICATIONS

Our study results support our primary hypotheses that: people naturally understand SBD and can use it to create robotic behaviors, our interfaces are usable and enable people to comfortably demonstrate interactive, stylistic robot locomotion behaviors, and the underlying Puppet Master algorithm generated behaviors that participants were able to differentiate, and showed potential for learning and mimicry results. In addition the study lends strong support for the idea of robots communicating style through their locomotion paths. Problems that emerged were either from interface issues (e.g., the small Surface) or from Puppet Master (e.g., movement jitter); no participant expressed issue with the SBD approach.

In addition to these primary results, we presented a descriptive qualitative evaluation that provides contextual insight into how people engage SBD. We also presented several substantial SBD-related findings that emerged from this study, and that we believe will be useful for helping direct future SBD work.

As one last point for reflection: we previously performed a similar animation-only study using a much earlier (non-robot) variant of the Puppet Master algorithm [Young et al., 2008], and offer a quick comparison here. Perhaps the most striking difference was that, while character movement jitter was also a problem for the animation Puppet Master version, animation participants did not use anthropomorphism or character traits to explain it. Unlike in the robot case, where the robot was assumed to be "confused," jittering animated characters were simply described as visually distracting and annoying. Further, unlike in this robot work, animation participants did not voice any concerns over the aggressive animated characters as they did with the *burglar* robots; we presume this is because animated actions do not have real-world implications in the same way that robotic actions do. These findings suggest that there is a fundamental difference in how movements are interpreted in the real world versus on a screen, and thus have important implications for animated work that simulates human-robot interaction.

8.1. Implications and Guidelines

Here we summarize our overall findings into design guidelines for future work:

Robotic style-by-demonstration is feasible – People understand SBD and do not require training to program style-oriented interactive robot behaviors by demonstration.

Robotic stylistic locomotion is feasible – People readily attribute meaning to a robot's locomotion style, and this can be used as an integral part of the robot's behavior, personality, and communication repertoire.

People Understand Teaching – People have an acute understanding of the challenges, difficulties, and nuances of teaching others, and can readily apply this knowledge to interaction with robots. People understand that teaching style relies on the robot's interpretation, and can adapt (such as by appropriate exaggeration or emphasis) to accommodate. We note that this does not necessarily mean the user is a *good* teacher.

Teaching a robot changes how people interact with it – The act of teaching a robot impacts how people perceive the robot's interaction. It can raise elements of responsibility for the robot's actions, and can increase acceptance of the robot's faults and limitations.

9. CONCLUSIONS

In this paper we detailed formal studies that explore people's interaction experiences with teaching interactive, stylistic behaviors to robots. We presented original style-by-demonstration interfaces and methods, specifically targeting the style of interactive locomotion, and thoroughly evaluated these through formal evaluation. We used exploratory qualitative methods to probe the SBD interaction experience, and presented a simple set of resulting design guidelines.

Overall, our results highlight how people readily accept and understand the ideas of teaching to robots, of teaching abstract *style*-oriented behaviors, and of robots that use stylistic behavior for communication. Thus, we have presented concrete ways that robots can leverage interaction skills that people already possess through *the social stock of knowledge* [Berger and Luckmann, 1966], to make complex HRI problems easy and accessible. We hope that this general direction, teaching style by demonstration, will continue to expand into new and exciting directions.

REFERENCES

[Amaya et al., 1996] Amaya, K., Bruderlin, A., and Calvert, T. (1996). Emotion from motion. In *Proc. 1996. GI '96, Toronto, Canada, May 22–24, 1996*, pages 222–229.

[Bartneck et al., 2007] Bartneck, C., Verbunt, M., Mubin, O., and Mahmud, A. (2007). To kill a mockingbird robot. In *Proc. HRI '07, Washington, D.C., Mar. 10–12, 2007*, pages 81–87, NY, USA. ACM.

[Belk, 1988] Belk, R. W. (1988). Possessions and the extended self. *Journal of Consumer Research*, 15(2):139–168.

[Berger and Luckmann, 1966] Berger, P. L. and Luckmann, T. (1966). *The social construction of reality: a treatsie in the sociology of knowledge*. Anchor Books, NY, USA.

[Bernard, 2000] Bernard, H. R. (2000). *Social Research Methods: Qualitative and Quantitative Approaches*. Sage Publications, Thousand Oaks, CA, USA.

[Breazeal, 2002] Breazeal, C. L. (2002). *Designing Sociable Robots*. The MIT Press, Cambridge, MA, USA.

[Breazeal et al., 2004] Breazeal, C. L., Brooks, A. G., Gray, J., Hoffman, G., Kidd, C. D., Lee, H., Lieberman, J., Lockerd, A., and Mulanda, D. (2004). Tutelage and collaboration for humanoid robots. *International Journal of Humanoid Robotics (IJHR)*, 1(2):315–348.

[Field, 2009] Field, A. (2009). *Discovering Statistics through SPSS: (and sex and drugs and rock 'n' roll)*. Sage, USA.

[Forlizzi and DiSalvo, 2006] Forlizzi, J. and DiSalvo, C. (2006). Service Robots in the Domestic Environment: A Study of the Roomba Vacuum in the Home. In *Proc. HRI '06, Salt Lake City, USA, Mar. 2–4, 2006*, pages 258–256, NY, USA. ACM.

[Frei et al., 2000] Frei, P., Su, V., Mikhak, B., and Ishii, H. (2000). Curlybot: designing a new class of computational toys. In *Proc. CHI '00, The Hague, The Netherlands, Apr. 1–6, 2000*, pages 129–136, NY, USA. ACM.

[Gallaher, 1992] Gallaher, P. E. (1992). Individual differences in nonverbal behavior: Dimensions of style. *Journal of Personality and Social Psychology*, 63(1):133–145.

[Gribovskaya and Billard, 2008] Gribovskaya, E. and Billard, A. D. (2008). Combining dynamical systems control and programming by demonstration for teaching discrete bimanual coordination tasks to a humanoid robot. In *Proc. HRI '08*, pages 33–40, NY. ACM.

[Guo et al., 2009] Guo, C., Young, J. E., and Sharlin, E. (2009). Touch and toys: new techniques for interaction with a remote group of robots. In *Proc. CHI, 2009. CHI '09, Boston, USA, April 4–9, 2009*, pages 491–500, NY, USA. ACM.

[Halbert, 1984] Halbert, D. (1984). *Programming by Example*. PhD thesis, University of California Berkeley.

[Harris and Sharlin, 2011] Harris, J. and Sharlin, E. (2011). Exploring the affect of abstract motion in social human-robot interaction. In *IEEE International Workshop on Robot and Human Interactive Communication, 2011. ROMAN '11, Atlanta, US, July 31-Aug 3, 2011*, NY, USA. ACM.

[Heider and Simmel, 1944] Heider, F. and Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, 57:243–259.

[Kanda et al., 2007] Kanda, T., Kamasima, M., Imai, M., Ono, T., Sakamoto, D., Ishiguro, H., and Anzai, Y. (2007). A humanoid robot that pretends to listen to route guidance from a human. *Auton. Robots*, 22(1):87–100.

[Kiesler and Hinds, 2004] Kiesler, S. and Hinds, P. (2004). Introduction to This Special Issue on Human-Robot Interaction. *Human Computer Interaction*, 19(1/2):1–8.

[Kiesler and Kiesler, 2004] Kiesler, T. and Kiesler, S. (2004). My pet rock and me: An experimental exploration of the self extension concept. *Journal of Consumer Research*, 32.

[Long et al., 2001] Long, A. C., Landay, J. A., and Rowe, L. A. (2001). "those look similar!" issues in automating gesture design advice. In *Proceedings of the Workshop on Perceptive User Interfaces, 2001. PUI '01, Orlando, Fl, USA, November 15–16*, pages 15–16, NY, USA. ACM.

[Matsui et al., 2005] Matsui, D., Minato, T., MacDorman, K. F., and Ishiguro, H. (2005). Generating Natural Motion in an Android by Mapping Human Motion. In *Proc. IROS, '05, Edmonton, Canada, Aug. 2–6, 2005*, pages 1089–1096, CA, USA. IEEE.

[McDaniel and Myers, 1999] McDaniel, R. G. and Myers, B. A. (1999). Getting more out of programming-by-demonstration. In *Proc. CHI, 1999. CHI '99, Pittsburgh, PA, USA, May 15–20, 1999*, pages 442–449, NY, USA. ACM.

[Raffle et al., 2004] Raffle, H. S., Parkes, A. J., and Ishii, H. (2004). Topobo: a constructive assembly system with kinetic memory. In *Proc. CHI '04, Vienna, Austria, Apr. 24–29, 2004*, pages 647–654, NY, USA. ACM.

[Reeves and Nass, 1996] Reeves, B. and Nass, C. (1996). *The Media Equation: How people treat computers, television, and new media like real people and places*. CSLI, Cambridge, UK.

[Saerbeck and Bartneck, 2010] Saerbeck, M. and Bartneck, C. (2010). Perception of affect elicited by robot motion. In *Proc. HRI ’10, Osaka, Japan, Mar. 2–5, 2010*, pages 53–60, NY, USA. ACM.

[Short et al., 2010] Short, E., Hart, J., Vu, M., and Scassellati, B. (2010). No fair!!: an interaction with a cheating robot. In *Proc. HRI ’10, Osaka, Japan, Mar. 2–5, 2010*, pages 219–226, NY, USA. ACM.

[Strauss and Corbin, 1998] Strauss, A. and Corbin, J. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications, London.

[Sung et al., 2009] Sung, J., Christensen, H. I., and Grinter, R. E. (2009). Robots in the wild: understanding long-term user. In *Proc. HRI ’09, San Diego, California, USA, Mar. 11–13, 2009*, pages 45–52, Berlin. ACM, Springer.

[Sung et al., 2007] Sung, J.-Y., Guo, L., Grinter, R. E., and Christensen, H. I. (2007). “my Roomba is Rambo”: Intimate home appliances. In *Proc. UBICOMP ’07, Innbruck, Austria, Sept. 16–17, 2007*, pages 145–162. Springer-Verlag, Berlin, New York, Heidelberg.

[Voida et al., 2011] Voida, A., Harmon, E., and Al-Ani, B. (2011). Homebrew databases: Complexities of everyday information management in nonprofit organizations. In *Proc. CHI, 2011. CHI ’11, Vancouver, Canada, May 7–12, 2011*, NY, USA. ACM.

[Wahlström et al., 2011] Wahlström, M., antti Salovaara, Salo, L., and Oulasvirta, A. (2011). Resolving safety-critical incidents in a rally control center. *Human Computer Interaction*, 26:9–37.

[Wobbrock et al., 2005] Wobbrock, J. O., Aung, H. H., Rothrock, B., and Myers, B. A. (2005). Maximizing the guessability of symbolic input. In *Conference Abstracts and Applications, Proc. CHI, 2005. CHI ’05, Portland, OR, USA, April 2–7, 2005*, pages 1869–1872, NY, USA. ACM.

[Wolber, 1997] Wolber, D. (1997). Pavlov: an interface builder for designing animated interfaces. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 4(4):347–386.

[Young et al., 2009] Young, J. E., Hawkins, R., Sharlin, E., and Igarashi, T. (2009). Toward acceptable domestic robots: Applying insights from social psychology. *Inagural Issue of the Int J Social Robot.*, 1(1):95–108.

[Young et al., 2008] Young, J. E., Igarashi, T., and Sharlin, E. (2008). Puppet master: Designing reactive character behavior by demonstration. In *ACM SIGGRAPH / Eurographics Symposium on Computer Animation, 2008. SCA ’08, Dublin, Ireland, July 7–9, 2008*, pages 183–191, Germany. Eurographics Association Press, Eurographics Association.

[Young et al., 2010a] Young, J. E., Ishii, K., Igarashi, T., and Sharlin, E. (2010a). Style-by-demonstration: Using broomsticks and tangibles to show robots how to follow people. In *adjunct proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (Late-Breaking Abstracts), 2010. HRI LBA '10, Osaka, Japan, March 2–5, 2010*, NY, USA. ACM.

[Young et al., 2010b] Young, J. E., Ishii, K., Igarashi, T., and Sharlin, E. (2010b). Style-by-demonstration: Using broomsticks and tangibles to show robots how to follow people. Departmental technical report 2010-975-24, University of Calgary, Calgary, AB, Canada.

[Young et al., 2010c] Young, J. E., Sung, J., Voida, A., Sharlin, E., Igarashi, T., Christensen, H., and Grinter, R. (2010c). Evaluating human-robot interaction: Focusing on the holistic interaction experience. *Int J Social Robot*, 3(1):53–67.