King's Research Portal

Document Version

Early version, also known as pre-print

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Sena, A., Zhao, Y., & Howard, M. J. W. (2018). Teaching Human Teachers to Teach Robot Learners. In *International Conference on Robotics and Automation (ICRA)* IEEE.

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

©2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

This is a pre-print document, which has since been accepted for publication in the 2018 IEEE International Conference on Robotics and Automation. This document will be updated with a full citation to the original IEEE publication, including DOI, when available.

Teaching Human Teachers to Teach Robot Learners*

Aran Sena^{1*}, Yuchen Zhao¹, and Matthew J. Howard¹

Abstract—Using Programming by Demonstration to teach robot learners generalisable skills relies on having effective human teachers. This paper aims to address two problems commonly observed in demonstration data sets that arise due to poor teaching strategies; *undemonstrated states* and *ambiguous demonstrations*. Overcoming these issues through the use of visual feedback and simple heuristic rules is investigated as a potential way of training novice users to more effectively teach robot learners to generalise a task. The proposed method intends to offer the user a more transparent understanding of the robot learner’s model state during the teaching phase, to create a more interactive and robust teaching process. Results from a single-factor, three-phase repeated measures study with $n = 30$ participants, comparing the proposed feedback and heuristic rules set against an unguided condition, show a statistically significant ($F(2, 58) = 8.0289, p = 0.00084$) improvement of user teaching efficiency of approximately 180% when using the proposed feedback visualisation.

I. INTRODUCTION

Programming by Demonstration (PbD) has long been explored as an approach for allowing robot operators to rapidly automate tasks, and with the rise of collaborative robots, PbD is becoming common in industrial settings. In addition to speed of deployment, PbD removes (or at least greatly reduces) the requirement for conventional programming skills to deploy robots, making robot programming accessible to *novice* users, *i.e.*, people who may not have a relevant technical skill set [1].

A basic use of PbD is direct demonstration of trajectories or trajectory way-points, either through kinesthetic teaching (physically guiding the robot through the desired movement) or teleoperation/jogging, followed by (repeated) playback of the demonstrated. However, a more scalable approach is to use a robot learner capable of *generalisation* from user demonstrations to accommodate possible variations in task parameters [2].

However, deploying systems that can learn generalised tasks from novice teachers presents a number of challenges. When a system incorporates human input to its learning process, the system performance heavily depends on the quality of the human-provided data. As noted in [3], two key data set issues that can arise as a result of poor teaching in PbD are *undemonstrated states*, and *ambiguous demonstrations* which might confuse the robot learner.

The contribution of this paper is to show that these problems can be mitigated through use of visual feedback and simple heuristic rules. It is shown that such methods offer the user a more transparent view on the robot learner’s

Fig. 1: Example of programming a machine tending task through PbD, by (a) providing task demonstrations for *every* position, and (b) providing only a few *key* demonstrations to a learner capable of generalisation. Example positions successfully reached by the learner during reproduction are marked in green and unsuccessful ones in red.

“understanding” of the task (*i.e.*, the current model state) during the teaching phase, creating a more interactive and robust process.

Results from a single-factor three-phase repeated measures study with $n = 30$ participants show a statistically significant improvement of user teaching efficiency of approximately 180% (as determined by the ratio of generalisation performance against the required number of demonstrations). This suggests the use of such measures is an important tool in improving system performance, and aiding novice teachers to teach robot learners.

II. BACKGROUND

A. Skill Generalisation in Collaborative Robots

Collaborative robots in industry have allowed novice users to automate many repetitive tasks using kinesthetic teaching and simplified programming interfaces for defining way-points, trajectories, and procedures [4], [5]. While this is useful for many basic tasks, programming more complex tasks can become time-consuming, as well as sensitive to variance in the task parameters (object positions *etc.*).

As an example, consider the robotic machine-tending task shown in Fig. 1. Here, the user must show the robot how to take gears from a conveyor (right) and place them onto one of the pegs on the parts-carrier (left). This task could be achieved through a traditional record-and-playback

¹Aran Sena, Yuchen Zhao and Matthew J. Howard are with the Center for Robotics Research, Department of Informatics, King’s College London, London, UK. aran.sena@kcl.ac.uk.

*This work was supported by the U.K. Agriculture and Horticulture Development Board (AHDB, HNS/PO14) and the Engineering and Physical Sciences Research Council (EPSRC) SoftSkills project, EP/P010202/1.

approach to PbD, if the user were to demonstrate the required trajectory for each peg, as in Fig. 1(a). However, such a laborious, time consuming approach is inefficient and non-robust.

Instead, it would be beneficial to be able to achieve the desired behaviour by providing just *a few, key demonstrations of the task*, and have the robot *generalise the behaviour* to the other peg locations. For instance, in such a scheme, the user might only have to demonstrate the trajectories shown in Fig. 1(b), from which the robot might generalise to the remaining pegs (red and green dots). Several learning methods capable of this kind of trajectory generalisation have been proposed in the literature, such as Dynamic Movement Primitives [6], Gaussian Mixture Models/Gaussian Mixture Regression [7], and approaches based on Hidden Markov Models [8].

Despite the advances in such generalised task learning methods, it is inevitable that failure cases may occur in which the robot fails to generalise correctly – either due to limitations in its learning capabilities or misuse by naïve users. For instance, in the example in Fig. 1(b), it may be that, due to the concentration of demonstrated trajectories around top and bottom rows of pegs, the robot only successfully generalises the skill to the points highlighted in green, and not those highlighted in red. There are several potential causes for this, including (i) the presence of *undemonstrated states*, where a lack of demonstration data impedes learning performance (e.g., centre region of Fig. 1(b)), or (ii) *ambiguous demonstrations*, whereby the user provides demonstrations that do not provide sufficient information to improve the model (e.g., if the user provided repeated demonstrations of the top-most trajectory in Fig. 1(b), this provides no new information to the model and fails to correct the execution of the task by the robot).

While such problems may be reduced by the continued improvement in learning algorithms, it is likely that the *training of users of collaborative robots* may also have a significant role to play in improving teaching efficiency.

B. Role of the Teacher

Instead of developing a more sophisticated model to handle undemonstrated states and ambiguous demonstrations, the human teacher’s natural adaptability can be leveraged to improve the system performance.

A challenge with addressing the above issues by relying on the teacher is that novice users can lack an accurate mental model of the robot’s knowledge, leading to system performance issues due to misjudging factors such as its learning rate, coverage and data support of the model, and confidence intervals. As highlighted by [9], there is a need to maintain transparency in the robots’ knowledge during teaching and to offer more effective methods for communicating the knowledge state of the robot to the human. By offering insight to the teacher of the robot learner’s current knowledge during training and assessment, it is hypothesised that ambiguous demonstrations and undemonstrated states can be avoided, to improve teaching efficiency.

Indication of the usefulness of visualisation of the robot model state to help novice teachers can be seen in [10], where

Fig. 2: Link between task sets. As more task demonstrations are provided to the system, $\hat{\pi}$ improves, resulting in $\hat{I}$ and $\hat{\beta}$ converging toward I and β , respectively.

visualisation for improving action plans is explored. There, the authors focus on transferring high-level plans from the user to the robot system, with the user interface visualising the effect of high-level actions and presenting the user with options on how to modify them, rather than addressing task demonstration and generalisation at trajectory level. Additionally, PbD is used in [10] for way-point demonstration, with actual trajectory generation done through conventional Inverse Kinematics, thus the generalisation in this study is concerned with generalisation of task plans. The approach proposed in this study instead improves teaching performance when the user wishes to generalise over low-level trajectories, through visualisation of the robot’s current learned model state.

III. PROBLEM DEFINITION

This paper investigates the use of visual feedback and simple heuristic rules as a way of training novice users to more effectively teach a robot learner to generalise a task. In order to help define the experiment, it is useful to formalise the teaching problem and the metric being evaluated.

A. Task Generalisation

Let $\mathcal{X}$ denote the *task set*—the set of all admissible states for a given task, which excludes regions occupied by obstacles, and regions defined to be out-of-bounds. Let $\mathcal{I} \subseteq \mathcal{X}$ denote the *initiation set* for a given task, which describes all admissible start states for the task execution, and $\beta \subseteq \mathcal{X}$ the *terminal set* which describes all admissible final states for the task execution.

The *task policy*, π , then generates a trajectory, $\mathcal{T}$, that can link $\mathcal{I}$ to β , while ensuring the generated trajectory stays within $\mathcal{X}$. The task is thus represented by the tuple $(\mathcal{I}, \pi, \beta)$; a representation drawn from [11]. Note that, a task is only achievable *iff* the trajectory generated by the policy is a subset of the end-effector task space of the robot, $\mathcal{O}$, i.e., if $\mathcal{T} \subseteq \mathcal{O}$, the robot can reach all points in the trajectory.

A *learned task policy*, $\hat{\pi}$, is generated from M user demonstrations. Each user provided demonstration, $\mathcal{D}_m$, contains T_m data points which are then concatenated to form a data set, $\mathbf{D}$, of N points, with $N = \sum_m T_m$ [12]. As the policy may be trained with incomplete or incorrect data, trajectories generated by $\hat{\pi}$ may not be valid or completely cover $\mathcal{I}$ or β .

The generalisation ability of $\hat{\pi}$ is then characterised by the sets $(\hat{\mathcal{I}}, \hat{\beta})|_{\hat{\pi}}$, i.e., the sets $\hat{\mathcal{I}}$ and $\hat{\beta}$ of initiation and terminal states that are successfully linked by trajectories generated with $\hat{\pi}$. Using $\hat{\pi}$, a set of R trajectories are generated, $\mathcal{T}$, each containing T_r data points. A generated trajectory is

Fig. 3: Experimental set up. (a) Participants are asked to guide a lightweight robot through the workspace from the start zone to the goal location. (b) End-effector of the robot indicating the positioning of the TrakStar sensor used for recording data.

considered successful if it meets the criteria $\mathcal{T}_{r,1} \subseteq \mathcal{I}$, $\mathcal{T}_{r,T_r} \subseteq \beta$, and $\mathcal{T}_r \subseteq \mathcal{X}$, where the subscripts 1 and T_r indicate the *first* and *last* data points in the trajectory respectively. The link between sets is illustrated in Fig. 2.

Note also that, initially, $\hat{\mathcal{I}}$ and $\hat{\beta}$ begin as null sets $\emptyset$, but as the teacher provides task demonstrations, the goal is that eventually they converge toward $\mathcal{I}$ and β , respectively.

B. Generalisation for Point-reaching

In this study, $\mathcal{X}$ is defined by a two dimensional bounding rectangle (20 cm by 30 cm), shown in Fig. 3(a), containing the start zone, $\mathcal{I}$ (a 20cm by 6cm rectangle), and the target, β , a point. The two obstacles (shown in the figure as shaded blocks) are *not* included in $\mathcal{X}$.

As β is a singular point goal, the issue of undemonstrated states is focused on achieving generalisation over $\mathcal{I}$. Here, $\mathcal{I}$ is constrained to a finite size, and discretised into a set of R test points, forming an *ideal* initiation set size. A data set with undemonstrated states can then be identified by $|\hat{\mathcal{I}}| < |\mathcal{I}|$ for $m = M$ (i.e., the initiation set from the learned policy is smaller than R after the last demonstration), and the issue of ambiguous demonstrations can be defined as $|\hat{\mathcal{I}}_m| \leq |\hat{\mathcal{I}}_{m-1}|$ (i.e., the initiation set size from the learned policy does not change, or reduces, between demonstrations).

For this experiment, the robot's progress in learning and generalising the task can then be simply represented as the ratio between the initiation set size from the learned policy and the ideal initiation set size. This ratio is then normalised by the number of demonstrations provided to define a teaching efficiency metric η :

$$\eta = \frac{1}{M} \frac{|\hat{\mathcal{I}}|}{|\mathcal{I}|} \quad (1)$$

C. Experimental Hypotheses

Following from §II-B and §III, the experimental hypotheses are chosen to test whether the PbD data set issues of undemonstrated states and ambiguous user demonstrations can be addressed by providing a transparent teaching process through a visualisation of the robot's knowledge. The key metric used for evaluating the success of the proposed approach is the teaching efficiency, η .

H_1 : Visualisation of the robot's learning progress results in a significant improvement in teaching efficiency, compared against a no-guidance teaching process.

H_2 : A heuristically guided teaching process, which utilises visual feedback, results in significantly improved teaching efficiency, compared against a no-guidance teaching process.

H_3 : A heuristically guided teaching process which uses visual feedback results in significantly improved teaching efficiency, versus a solely visually guided teaching process.

The first two hypotheses, H_1 and H_2 , are designed to test a novice user's ability to cover all initial states $\mathcal{I}$. H_1 evaluates the impact of purely visual feedback system. H_2 introduces a heuristic rule set for teacher guidance. H_3 is designed to check for differences between the two visual feedback methods used in H_1 and H_2 .

IV. MATERIALS & METHODS

This section describes experimental procedures in testing the above hypotheses with human subjects¹ and details of the analysis conducted on the experimental data².

A. Teaching Task

To test hypotheses H_1 – H_3 , a simple experiment is used in which participants are asked to teach a lightweight robotic arm how to move its end-effector from a “start-zone” to a point target in a fixed environment, containing obstacles, using kinesthetic PbD.

The participants are asked to grip a lightweight robot by its end-effector and guide it through the two-dimensional workspace while avoiding the two shaded obstacles, shown in Fig. 3(a). As mentioned in §III, the workspace is restricted to the two-dimensional XY plane of the table top, resulting in demonstrations $\mathcal{D}_m = \{r_t^x, r_t^y\}_{t=1}^{T_m}$, where $\mathbf{r}_t \in \mathbb{R}^2$ is the robot end-effector position at sample t . This is ensured by asking participants to maintain contact between the robot end-effector and the work surface at all times. Demonstrations and trajectories which leave $\mathcal{X}$ are considered to have failed the task. This task is suitable for testing the hypotheses as (i) task complexity is restrained so as not to have an adverse affect on teaching quality due to user skill, and (ii) it requires effective demonstration generalisation in order to successfully cover the start zone (i.e., $\mathcal{I}$).

The robot used is a uFactory uArm Metal, a lightweight robot (< 1 kg) with back-driveable motors. The low inertia and backdrivability of the robot means there is minimal interference with the user demonstrations due to the dynamics of the robot. As demonstration data for the task parametrised learning methods (ref. §IV-C), the end-effector position of the robot is recorded at 80Hz as participants guide the robot through the task. It should be noted that, due to mechanical play in the joints, computation of the end-effector position through forward kinematics based on joint-encoder data has been found to be unreliable for this robot. To address this, an NDI TrakStar sensor is placed co-axial with the end-effector

¹This experiment was approved by a KCL ethics committee, ref. LRS-16/17-3800. Informed consent was obtained from all experimental participants.

²The data supporting this research will be made open access with accreditation upon publication of the paper. Further information about the data and conditions of access can be found by emailing research.data@kcl.ac.uk

Fig. 4: Example of a participant provided demonstration sequence where $M = 5$, and the visualisation shown to participants during the visual feedback test conditions. m is number of demonstrations given at the presented snapshot.

tip, providing $\pm 1.3 \text{ mm}$ RMSE positioning accuracy, shown in Fig. 3(b).

B. Training Procedure

As the experimental participants are novice users with little or no background in robotics, all participants are provided with a basic level of training. This first involves an introductory video, which explains PbD through the same machine tending task example as described in §II-A. The participants are then shown a series of short, instructional videos, staged throughout the experiment³. Videos are used for experiment instruction to help ensure consistency in the participants' prior knowledge for each teaching phase, and to remove the risk of the researcher providing more or less information on how to complete the task between participants. Additionally, as noted in [13], video instruction is a highly effective teaching tool for novice users.

After the introductory video, the participant is given one minute to familiarise themselves with the robot where they are free to move it around. After this, the participant takes part in the three phases of the experiment.

Teaching Phase 1 - No Guidance (NG): In this phase of teaching, the participant is asked to provide task demonstrations so that the robot will be able to generate successful trajectories to the goal from anywhere in the start zone. This phase is conducted to test whether participants are able to provide demonstrations with no feedback and little prior knowledge, and avoid giving ambiguous demonstrations or having undemonstrated states at the end of the teaching phase. The instruction video explains the different areas of the task map, provides one basic example of how the task is meant to be performed, and explains that they must provide as many demonstrations as they feel are necessary so that the robot can perform the task from anywhere inside the start zone.

Teaching Phase 2 - Visual Feedback, No Guidance (VFNG): In this phase, the effect of providing a transparent visualisation of the robot's learning progress is tested. In between each demonstration, the user is provided with a visualisation of the learning progress, as shown in Fig. 4, which highlights successful start locations and trajectories in green, and failed points in red. The instruction video explains this visualisation, and provides a simple example of what demonstrations look like in the visualisation. The video explains they must continue to provide demonstrations until

they have reached full coverage of the start zone (all test locations highlighted green). The division of the start zone for trajectory generation is further discussed in §IV-D.

Teaching Phase 3 - Visual Feedback, Rule Guidance (VFRG): In this phase the participant must now follow a set of rules when providing their demonstrations. The rule set is designed to approximately guide users towards providing demonstrations which maximise the likelihood of increasing the information gain for each demonstration.

The rules are: (i) provide one demonstration, starting from anywhere in the start zone, (ii) continue providing demonstrations within 4cm of the first demonstration, until the first demonstration is surrounded by successful test points, (iii) provide further demonstrations within 4 cm of the successful test points, in the area with the greatest number of failed test points.

The instruction video explains the rules through an example which takes steps to avoid influencing the participants' behaviour during teaching, such as only showing a small section of the start zone when explaining the rule set.

After each demonstration, the robot learns a policy from the data set $\mathbf{D}_1 = \{r_s^x, r_s^y\}_{s=1}^N$, where $\mathbf{r}_s \in \mathbb{R}^2$ is the position of the robot end-effector for data set sample s . This learned policy is only visualised in the VFNG and VFRG conditions. In the NG condition, $\hat{\pi}$ is only used for teaching efficiency evaluation at the end of the test phase.

C. Task Learning

To examine the effect of the different teacher training schemes outlined in §IV-B on teaching efficiency, the data from the robot is provided to a task parametrised learning algorithm to learn a generalised model of the task. In this paper, the task parametrised learning method used is Task-Parametrised Gaussian Mixture Model (TP-GMM), combined with Gaussian Mixture Regression (GMR). The latter has been shown to be effective in learning task-parametrised generative models [7], [14] and allows the generation of new trajectories from the learnt encoding at low computational cost [12].

TP-GMM works by learning a task from multiple frames of reference, which are combined through a linear transformation to give a single representation of the task. Each frame is defined by an offset vector $\mathbf{b}_{m,p} \in \mathbb{R}^2$, and a linear transformation matrix $\mathbf{A}_{m,p} \in \mathbb{R}^{3 \times 3}$ where the subscript m is the demonstration index, and p is the frame index ($p = 1$ for the initiation frame, $p = 2$ for the termination frame). In the experiments reported here, four frames of reference are used in the model: two frames are fixed obstacle frames

³All videos used in the experiment are provided as supplementary data to this paper, and are available to view online <https://www.youtube.com/watch?v=19VgW0dsbY>.

located at the centre of each obstacle, while the other two are the initiation state frame and terminating state frame extracted from the demonstration data.

For each demonstration, the initiation state frame offset vector is taken as the first data point in the demonstration, $\mathbf{b}_{m,p} := \mathbf{r}_{m,1}$ with $p = 1$ for the initiation frame. Similarly, the termination state frame is taken from the last data point, $\mathbf{b}_{m,p} := \mathbf{r}_{m,T_m}$ with $p = 2$ for the termination frame.

Each $\mathbf{A}_{m,p}$ is then estimated from the demonstration trajectory:

$$\mathbf{A}_{m,p} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & (b_{m,p}^y - x_{m,j}^y) & (x_{m,j}^x - b_{m,p}^x) \\ 0 & (x_{m,j}^x - b_{m,p}^x) & (x_{m,j}^y - b_{m,p}^y) \end{bmatrix} \in \mathbb{R}^{3 \times 3} \quad (2)$$

where, the x/y raised index is used to indicate the component of the vector to use, and $\mathbf{x}_{m,j} \in \mathbb{R}^2$ is the j^{th} end effector position from the m^{th} demonstration that will result in a frame length of at least l^* from $\mathbf{b}_{m,p}$, as determined by the Euclidean distance:

$$\mathbf{x}_{m,j} = \underset{\mathbf{x}_{m,i}}{\operatorname{argmin}} \left\| \sqrt{(\mathbf{b}_{m,p} - \mathbf{x}_{m,i})^T (\mathbf{b}_{m,p} - \mathbf{x}_{m,i})} - l^* \right\|^2 \quad (3)$$

where $\mathbf{x}_{m,i}$ is the i^{th} data point in the m^{th} demonstration, and for this experiment $l^* = 1.0\text{cm}$.

A model parameter to be selected by the researcher is the number of Gaussians the TP-GMM should use to encode the task trajectory. Based on preliminary tests a mixture of 11 Gaussians provided reasonable reproduction behaviour.

Once the required frames have been estimated, these are provided to the TP-GMM model along with the data set to learn the policy $\hat{\pi}$. This process for learning is repeated after each demonstration and used for evaluation or visualisation as described in §IV-B, but is reset at the end of each teaching phase so that the learned policy is unique for each phase.

D. Trajectory Generation

In order to test how well the robot can generalise the task from the provided demonstrations, a set of test trajectories are generated through forward-simulation for a given $\mathcal{I}$ and β . These are evaluated to see how many trajectories successfully link states in $\mathcal{I}$ to states in β while staying in $\mathcal{X}$. These trajectories are then also used for the visualisation provided to the user.

In this experiment, we consider the constrained case, for a task with a discrete $\mathcal{I}$, and a single state for β . $\mathcal{I}$ is divided into a test grid with R states, with each state represented by an offset vector and a transformation matrix, $\{\mathbf{b}_{r,p}, \hat{\mathbf{A}}_{r,p}\}_{r=1}^R$. In this experiment $R = 140$, representing a 20×7 test grid of points separated by 1cm.

The robot must use the learned policy, $\hat{\pi}$, to reproduce trajectories which attempt to reach β for any member of $\mathcal{I}$.

As mentioned in §IV-C, we require an offset vector $\mathbf{b}$ and transformation matrix $\mathbf{A}$ for each test point in $\mathcal{I}$. An estimate transformation matrix, $\hat{\mathbf{A}}_{r,p}$ for each of these test points is given by a weighted value of all the initiation frame $\mathbf{A}$ values in the test phase data set using an Inverse Distance Weighting function with a Gaussian Radial Basis Function (GRBF) weighting:

$$\hat{\mathbf{A}}_{r,p} = \frac{1}{Z} \sum_{m=1}^M w_m \mathbf{A}_{m,p} \in \mathbb{R}^{3 \times 3} \quad (4)$$

where $Z = \sum_{m=1}^M w_m$ and $w_m = e^{-\mathbf{z}^\top \mathbf{z} / 2\sigma^2}$ with $\mathbf{z} = \hat{\mathbf{b}}_{r,p} - \mathbf{b}_{m,p}$, and $p = 1$ for the initiation frame.

Trajectories can fail either due to model-based failure conditions, where the trajectory does not start in $\mathcal{I}$ and/or does not end in β , or task-based failure conditions, where the trajectory enters an out-of-bounds area (e.g., passes through an obstacle, or leaves the task area).

E. Statistical Analysis

As this experiment is designed as a within-subjects study, a repeated-measures Analysis of Variance (ANOVA) study is used to determine if any significant effects can be observed in the data.

The independent variable across the experiment phases is the level of guidance provided to the participant. In phase one, this is the no-guidance (NG) condition. In phase two, the participant is provided with visual feedback, but no other guidance (VFNG). In phase three, the participant is provided with both the visual feedback as well as a heuristic rule guidance (VFRG). In all stages of the experiment, the dependent variable is the participant's teaching efficiency, η .

A power analysis conducted using *GPower 3.1* [15] for a repeated measures ANOVA with one group and three measurement levels indicates that for a medium effect size (Cohen's $f = 0.3$), and a power of 0.95, the required sample size is $n = 30$.

The number of demonstrations used for evaluating user performance was taken to be the number of demonstrations required to achieve at least 90% coverage of the test grid, or the maximum number of demonstrations if 90% coverage was not achieved. 90% coverage is used for evaluation, as it was found in preliminary tests that by limiting demonstration start locations to points within the test grid, achieving full performance for the edge test locations was difficult to achieve.

Before performing analysis with repeated measures ANOVA, there are five underlying assumptions that must be checked. (i) *Continuous dependent variable*: The teaching efficiency is measured on a continuous scale $[0, 1] \in \mathbb{R}$, satisfying the first assumption. (ii) *At least two groups of the independent variable*: The independent variable, the feedback method, has three groups - No Guidance, Visual Feedback/No Guidance, and Visual Feedback/Rule Guidance. (iii) *Absence of outliers*: The data is checked for outliers. (iv) *Dependent variable normally distributed*: Using an Anderson-Darling test, the data for each test condition was tested for normality, with each test giving a $p > 0.05$, indicating that the null hypothesis (the data *has* come from a normal distribution) cannot be rejected. (v) *Assumption of sphericity*: Using a Mauchly's test for sphericity on the repeated measures model gives $p = 0.1233$, indicating no data correction is required.

V. RESULTS

The experiment was conducted with 30 participants (17 male, 13 female; mean age 35.1, SD=9.7). Analysing the teaching efficiency score for achieving 90% coverage of the test grid, or the teaching efficiency at the end of the participant demonstrations if 90% coverage was not achieved, the

Fig. 5: Box plot of results for each test condition. No Guidance (NG), Visual Feedback/No Guidance (VFNG), Visual Feedback/Rule Guidance (VFRG)

Fig. 6: Generalisation score versus number of demonstrations for the three test conditions. The curves represent the mean score of all participants at a given number of demonstrations, for each test condition. 90% performance threshold shown with a dashed line.

data indicates a significant effect of the independent variable, feedback method, on the dependent variable, teaching efficiency; $F(2, 58) = 8.0289, p = 0.00084$.

As there was a significant effect observed, a multiple comparisons of means was performed. A significant difference was found between the No Guidance (NG) condition and the visual feedback with no guidance (VFNG), $p = 0.0058$, as well as between the NG condition and the visual feedback with rules guidance (VFRG), $p = 0.0166$. No significant difference was observed between the VFNG condition, and VFRG condition, $p = 0.8008$.

VI. DISCUSSION

The results show a clear benefit in providing a visualised feedback to the teacher during PbD, with the repeated measures ANOVA indicating significance in the result; $F(2, 58) = 8.0289, p = 0.00084$.

Fig. 5 shows the mean and the spread of the teaching efficiency across the different conditions. As can be seen, performance both improves and becomes more consistent amongst participants with an increase in mean teaching efficiency and a reduction in the standard deviation in the VFNG and VFRG cases compared to the NG case.

Fig. 6 shows how the mean *generalisation score* across all participants varies as the number of demonstrations provided increases. When no feedback is provided (NG condition) participants tend to underestimate the number of demonstrations required to cover the initiation state space. This can be seen

in the red line in Fig. 6 which shows that the majority of participants left undemonstrated states in the NG condition (generalisation score under the dashed threshold line). Even for participants who provide many demonstrations, this is not guaranteed to avoid undemonstrated states as they may provide multiple ambiguous demonstrations from similar locations which does not improve the robot’s knowledge. While there were some participants who did achieve high teaching efficiency in the NG condition (7 participants), it was difficult for the majority. Conversely, in the VFNG and VFRG, all participants reached the threshold generalisation score, indicating that the visualisation helped to address the issue of undemonstrated states for novice teachers.

The results therefore show the first two hypotheses, H_1 and H_2 , can be accepted. Observing the average difference in medians between NG (0.0757) and the VFNG (0.1342) and VFRG (0.1398) conditions, it can be seen that the average teaching efficiency improves by approximately 180%. These results indicate that both visual feedback methods are able to address the issue of undemonstrated states and ambiguous demonstrations, compared against the no guidance condition.

No significant difference can be observed between the VFNG and VFRG conditions ($p = 0.8008$), so the H_3 must be rejected. Despite this, some interesting trends can be observed in the data. Looking at the spread of results in the box plot shown in Fig. 5, it can be seen that rule-based guidance, VFRG, appears to have smaller variance compared to the VFNG condition, and a higher mean. This could be indicative of an improvement in teaching performance when the participant is provided with a heuristic rule set. However, given the within-subjects design of the experiment there is a possibility that this is a learning effect artefact as the participants may develop a more accurate mental model for how the robot learns after observing the visualisation in Teaching Phase 2.⁴

In addition to differences in variance, it can be seen in Fig. 7 that the heuristic rule set did influence the behaviour of the participants, if we compare the NG and VFNG cases against the VFRG case. Fig. 7 depicts the relative probability (colour map) of a user starting in a given area of the start zone (vertical axis of Fig. 7) for successive demonstrations (left to right). For example, in Fig. 7(a) and (b), for the first demonstration ($k=1$) the user is most likely to give a demonstration in the bottom or top region of the start zone, and for the last demonstration in Fig. 7(a) ($k=5$) they are most likely to start in the middle of the start zone.

By observing the trends for where the participants start their demonstrations from, two key insights can be found. First, the NG and VFNG conditions show users are most likely to select starting points at the extremes of the start zone. Secondly, in the VFRG condition, it can clearly be seen that there is a shift in demonstration behaviour, with users tending to prefer starting in the bottom of the start zone, moving their demonstrations toward the top of the start zone with successive demonstrations. This seems to be in agreement with related work from the Active Learning

⁴The teaching phases order could not be randomised for this experiment, as it was important to capture the participants’ natural teaching style in Teaching Phase 1 and 2, before introducing a structured process in phase 3.

Fig. 7: Start locations for first K demonstrations from all participants for each test case, where K is the mean number of demonstrations for each test case. Each column represents the state of the test grid test after the n^{th} demonstration. Colourbar scale indicates probability of choosing a given vertical position.

domain, in which participants have been seen to favour teaching strategies that use “extreme” examples to convey the target concept (see [16], [17]).

As the performance results for VFNG and VFRG are similar, despite the differences in strategy, this suggests the possibility of redesigning teaching strategies in certain contexts. For example, VFRG potentially lends itself to more complex initiation spaces $\mathcal{I}$ where the extreme boundaries of where the robot must operate may be harder to define for the user. If the user can rely on employing a set of local heuristic rules that helps them explore $\mathcal{I}$, and can use the generative visualisation method presented here, it may help them teach tasks in unconstrained $\mathcal{I}$ and β spaces more effectively.

VII. CONCLUSION

Understanding how novice users interact with robot learners is critical to understanding how to achieve the best performance from deployed systems. Considering that the main purpose of a human teacher during PbD is to provide an informative data set from which the robot can learn a task, further work is required in addressing both the issue of acquiring good quality demonstrations, and addressing undemonstrated states which may adversely affect system performance and generalisation.

Based on the findings of this research, it has been shown that human teachers struggle to provide demonstrations which robustly cover a task input space, $\mathcal{I}$. This is often caused by not understanding how many demonstrations are actually required for the underlying model to learn the task, and/or not being able to reliably identify undemonstrated states in $\mathcal{I}$. The results show that by providing visual feedback of the learning system progress, the human teacher is able to reliably provide enough informative demonstrations to achieve good generalisation performance in the given task. While the results do not show a performance gain when using the heuristic rule set versus just the visual feedback, analysis of the participants’ demonstration behaviour indicates a rule set may be beneficial in future work where users must explore unconstrained task spaces.

Future work will also further explore the use of this training, evaluation and guidance method for unconstrained task spaces, higher complexity tasks, and the potential for heuristic guidance to improve PbD through an active learning process.

REFERENCES

- [1] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, “Robot Programming by Demonstration,” in *Springer Handbook of Robotics*. Berlin, Heidelberg: Springer, 2008, pp. 1371–1394.
- [2] S. Calinon and A. Billard, “What is the Teacher’s Role in Robot Programming by Demonstration? - Toward Benchmarks for Improved Learning,” *Interaction Studies. Special Issue on Psychological Benchmarks in Human-Robot Interaction*, vol. 8, no. 3, 2007.
- [3] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, “A survey of robot learning from demonstration,” *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469–483, 2009.
- [4] Rethink Robotics, “Intera 5 Online User Guide,” 2017. [Online]. Available: <http://mfg.rethinkrobotics.com/intera/Main-Page>[Accessed:2017-09-01]
- [5] Universal Robots, “Polyscope User Guide,” 2017. [Online]. Available: <https://www.universal-robots.com/download/?option=27420>[Accessed:2017-09-01]
- [6] S. Schaal, “Dynamic Movement Primitives -A Framework for Motor Control in Humans and Humanoid Robotics,” in *Adaptive Motion of Animals and Machines*. Tokyo: Springer-Verlag, 2006, pp. 261–280.
- [7] S. Calinon, F. Guenter, and A. Billard, “On Learning, Representing, and Generalizing a Task in a Humanoid Robot,” *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 2, pp. 286–298, 4 2007.
- [8] S. Calinon and A. Billard, “Recognition and reproduction of gestures using a probabilistic framework combining PCA, ICA and HMM,” in *International Conference on Machine Learning*. New York, USA: ACM Press, 2005, pp. 105–112.
- [9] R. Toris, H. B. Suay, and S. Chernova, “A practical comparison of three robot learning from demonstration algorithms,” in *International Conference on Human-Robot Interaction*. New York, USA: ACM Press, 2012, p. 261.
- [10] S. Alexandrova, M. Cakmak, K. Hsiao, and L. Takayama, “Robot Programming by Demonstration with Interactive Action Visualizations,” in *Robotics: Science and Systems X*, 7 2014.
- [11] M. Cakmak and A. L. Thomaz, “Active Learning with Mixed Query Types in Learning from Demonstration,” in *Proceedings of the ICML Workshop on New Developments in Imitation Learning*, 2011.
- [12] S. Calinon, “A tutorial on task-parameterized movement learning and retrieval,” *Intelligent Service Robotics*, vol. 9, no. 1, pp. 1–29, 9 2015.
- [13] M. Cakmak and A. L. Thomaz, “Eliciting good teaching from humans for machine learners,” *Artificial Intelligence*, vol. 217, pp. 198–215, 2014.
- [14] A. Pervez and D. Lee, “Learning task-parameterized dynamic movement primitives using mixture of GMMs,” *Intelligent Service Robotics*, pp. 1–18, 7 2017.
- [15] F. Faul, E. Erdfelder, A. Buchner, and A. Lang, “Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses,” *Behavior Research Methods*, vol. 41, pp. 1149–1160, 2009. [Online]. Available: <http://www.gpower.hhu.de/en.html>
- [16] M. Cakmak and A. L. Thomaz, “Optimality of human teachers for robot learners,” in *International Conference on Development and Learning*. IEEE, 8 2010, pp. 64–69.
- [17] F. Khan, B. Mutlu, and X. Zhu, “How Do Humans Teach: On Curriculum Learning and Teaching Dimension,” in *Neural Information Processing Systems*, 2011, pp. 1449–1457.