Second Language Tutoring using Social Robots

Project No. 688014

L2TOR
Second Language Tutoring using Social Robots

Grant Agreement Type: Collaborative Project
Grant Agreement Number: 688014

D4.4 State-of-the-art in robotic sensing for education technology

Due Date: **31/12/2018**
Submission Date: **20/01/2019**

Start date of project: **01/01/2016** Duration: **36 months**

Organisation name of lead contractor for this deliverable: **Plymouth University**

Responsible Person: **Tony Belpaeme** Revision: **1.0**

Project co-funded by the European Commission within the H2020 Framework Programme		
Dissemination Level		
PU	Public	**PU**
PP	Restricted to other programme participants (including the Commission Service)	
RE	Restricted to a group specified by the consortium (including the Commission Service)	
CO	Confidential, only for members of the consortium (including the Commission Service)	

Contents

Executive Summary

This deliverable is the final deliverable of WP4 (Multimodal input processing). It reports on the work of the last 9 months and offers a perspective on the state-of-the-art in processing input in Child-Robot Interaction. We look at the potential of Deep Learning for social signal processing, and looks at how multi-modal perception in robots can be used to build more robust sensor interpretation and consequently a more contingent interaction.

Principal Contributors

The main authors of this deliverable are as follows (in alphabetical order):

Tony Belpaeme, Madeleine Bartlett, Christopher Wallbridge, Emmanuel Senft (Plymouth University)
Bahar Irfan (Softbank Robotics Europe)

Revision History

Version 1.0 (T.B. 20-01-2018)
First and final version.

1 Deliverable Outline

This deliverable briefly surveys the state-of-the-art, going over the various tasks in the WP and input modalities needed for Human-Robot Interaction (HRI), and Child-Robot Interaction (CRI) in specific. It looks at evolutions in technology and assesses whether in three years since the start of L2TOR significant progress occurred in the field which are noteworthy for HRI and CRI.

1.1 Automated Speech Recognition (T4.1)

In the first year of the project, we noted during a rigorous evaluation study that both open and commercial Automated Speech Recognition solutions performed sub-optimally for children's speech. Especially in the context of L2TOR, where we wished to recognise speech from young children (4 to 6 year's old) all solutions showed nigh unusable recognition results [1]. This contrasted very much with the communication by vendors of speech recognition and with earlier reports from research [2, 3, 4].

In recent years there has been increased attention for child speech recognition. Driven by edutainment applications, such as interfaces for apps which do not rely on written language, there has been a drive to improve speech recognition for young children. YouTube Kids is perhaps the leading product on the market, but an informal evaluation showed that while the Google child speech recognition has improved due to more data being available (collected through the YouTube Kids app), the performance is still very much below what would be needed for successful HRI (see table 1).

Ground truth	Transcription by YouTube Kids
"one"	one
"two" (noisy)	*not recognised*
"three" (noisy)	*not recognised*
"the boy looking at the frog"	a boy ok
"and he saw a grasshopper on top of it"	installing a grass on top of it
"and a rat came euh popped out of the hole"	and wet K-pop telkomsel

Table 1: YouTube Kids, with Google ASR, transcribing a selection of speech utterances spoken by 5-year old native English speaking children. In early 2019, the performance of the ASR remains disappointing for utterances by young children.

Towards the end of 2018 a US start-up claimed to have robust child speech recognition. KidSense.ai has an edge solution, where children's utterances are transcribed on-board a device (as opposed to Google's cloud based solution). KidSense.ai has, despite advertising the availibility of a trial version of the software, not made their ASR engine available for evaluation. In the light of further evidence, we have to conclude that child speech recognition still is not sufficiently robust and cannot be used for spoken interactions with young children. The interaction, as such, will still need to run predominantly through the tablet interface.

1.2 Face detection and recognition (T4.2)

In recent years considerable progress has been made in the area of face detection and recognition, mainly driven by the use of Convolutional Neural Networks. While earlier feature-based solutions for face detection, such as the Viola-Jones face detection method [5], often under-performed on children due to their different physiognomy and dynamic behaviour in front of the camera, the Deep Learning based solutions do not seem to suffer from this. OpenCV, an open library for computer vision, starting

from version 3.3, which was released in mid 2017 contains a Deep Neural Network implementation for face detection which has proven to be effective for children.

In terms of face recognition, OpenFace, an open source implementation of FaceNet [6], implements face recognition using DNNs, and while the performance of OpenFace is reported to between 95% and 99% on benchmark databases, the performance in real world applications is much lower. A recent evaluation by partner PLYM showed that in real-world interactions with robots, the face recognition is very context-dependent suggesting that the high performance reported by FacedNet and similar solutions is due to overspecification on the benchmarks.

However, the fact that the interaction between the user (in our case a young child) and the robot is a multimodal interaction offers opportunities for more robust user identification. We studied how input from other modalities can be combined using a Bayesian Network to arrive at a effective user recognition [7]. The method uses information on the user's perceived gender, height, time of day of the interaction to improve user recognition. Our method increased the identification rate substantially up to 40% on open-set and closed-set scenarios.

1.3 Head pose, gaze tracking and gesture (T4.3)

Head pose and gesture tracking relied for a number of studies in L2TOR still on the Kinect SDK by Microsoft, which takes a feed from a Microsoft Kinect RGDB camera. However, since 2018, OpenPose has revolutionised the tracking of skeletal data (including head pose) through using a cascading DNN to track a total of 135 keypoints on single images. While the data returned by OpenPose is still a 2D coordinate, other solution exist to conver the 2D to a 3D data point. Given the robustness of OpenPose, the ability to deal with occlusion and highly dynamic scenes, and the need for only a single camera, it is expected that this approach, and similar ones, will substitute the use of RGBD cameras in HRI. This was used in the analyses of the PinSoRo dataset [8], a dataset of recording of more than 45 hours of interactions between children and children and robots, recorded and annotated by PLYM and made public, which serves to machine learn the extraction of interaction parameters.

Tracking eye gaze has also become relatively robust, and eye gaze trackers based on DNNs exist. One such eye tracker was developed and evaluated in the context of the FP7 DREAM project [9].

1.4 Emotion and affect recognition (T4.4)

Emotion recognition, while helped by recent evolutions in machine learning, still is rather primitive. The main two obstacles are on the one hand the limited categorisation of emotion, with Ekman's six basic emotions still (erroneously) considered as the only emotions worth recognising [10]. On the other hand, there is insufficient data which has been annotated with a richer coding scheme to train machine learning approaches to recognise and classify a richer set of emotions and affect.

Relevant to robots for learning is the recognition of "engagement", the rather difficult to define notion of the learner being mindful about the tutorig or learning experience. In L2TOR we did not use automated recognition of engagement, but instead relied on video coding by annotators to mark whether or not a child was engaged during the interaction. A number of projects rely on proxies, such as touch events or slowing down of the interaction, to gauge engagement [11].

In a recent study [12] we studied how internal states could be read from external signals, such as skeletal dynamics. We relied on Conceptors, a novel approach to classification with DNNs which does not only report classes, but which can arrive as a continuous assessment in between classes [13].

1.5 Tablet input (T4.5)

As natural language interaction proved difficult to implement, we relied to a large extent on the interaction with a touch screen tablet. This has proven to be reliable and accessible method of gathering input for the robot, and one that comes naturally to young children. Children of 4 or over do not need an introduction to the use of a tablet, with tapping and dragging well established when they first meet the robot.

Our expectations are that a touch screen will be the interface of choice for robot tutors for now and the foreseeable future.

1.6 Environment processing T(4.6)

Environment processing, the recognition of objects and events in the immediate vicinity of the child and the robot, has been made considerably easier by concentrating the offer of educational content on the tablet. An early study showed that there was no learning gain when objects were physically presented to the children when learning L2 words [14] and a decision was made to have all educational material displayed on the tablet. This not only made the identification and reading of the pose of objects considerably easier, but also allowed for the presentation of animated sequences, for example to teach the children verbs such as “running” and “climbing”.

Other elements relevant to environment processing include Voice Activity Detection, for which we relied on OpenSmile [15], and simple visual perception, for example to detect whether something is taking place in front of the robot.

2 Outlook

In terms of input processing for child-robot interaction, significant hurdles remain. While impressive progress has been made on interpreting the visual modality, the lack of performance in transcribing spoken language from young children forms a significant limitation for robots for language learning. While these limitations can be circumvented by using a different interface, such as a touch screen tablet, the initial promise of having a robot with which children could have a conversation remains unfulfilled and is likely to remain so for the foreseeable future.

A Annex Descriptions

A.1 Irfan et al. (2019) ulti-modal Incremental Bayesian Network with Online Learning for Open World User Identification

Bibliography - Irfan, B., Garcia Ortiz, M., Lyubova, N. and Belpaeme, T. (2019) Multi-modal Incremental Bayesian Network with Online Learning for Open World User Identification. *Submitted to Frontiers in AI and Robotics.*

Abstract - User identification is an important step in creating a personalised long-term interaction with robots, such as in domestic applications, education, or rehabilitation. It requires learning the users continuously and incrementally, possibly starting from a state without any known user. In this paper, we describe a multi-modal incremental Bayesian network with online learning to be applied in such scenarios. Face recognition is used as the primary biometric, and it is combined with ancillary information, such as gender, age, height and time of interaction obtained from proprietary algorithms, with a hybrid normalisation approach that combines the optimal normalisation method for each parameter to improve the recognition. We introduce the long-term recognition performance loss that weighs the importance of correct estimations of known users to the incorrect estimations of unknown users for optimising the parameters of the network. We generated multi-modal datasets with 200 users with random or periodic interaction times, that simulates an HRI scenario to evaluate our approach. The results show that the proposed network decreases the loss compared to face recognition alone and increases the identification rate substantially up to 40% in open-set and closed-set scenarios.

Relation to WP - This work contributes to Tasks T4.2 and T4.3.

A.2 Senft et al. (2019) Teaching robots social autonomy from in-situ human guidance

Bibliography - Senft, E., Lemaignan, S., Baxter, P., Bartlett, M. and Belpaeme, T. (2019) Teaching robots social autonomy from in-situ human guidance. *Submitted to Science Robotics*.

Abstract - Striking the right balance between human control and robot autonomy is a core challenge in social robotics, both in technical and ethical terms. On the one hand, extended robot autonomy offers the potential for increased human productivity and the off-loading of physical and cognitive tasks; on the other hand making the most of human technical and social expertise, as well as maintaining accountability, is highly desirable. This issue is particularly sensitive in domains such as medical therapy and education where social robots hold substantial promise, but where there is a high cost to poorly performing autonomous systems, compounded with sensitive ethical concerns. We present an ecologically valid study evaluating SPARC, a novel approach addressing this challenge whereby a robot progressively learns appropriate autonomous behaviour from in-situ human demonstrations and guidance. Using online machine learning techniques, we demonstrate that the robot can effectively acquire legible and congruent social policies in a high-dimensional child tutoring situation, from a limited number of demonstrations. By exploiting human expertise, our technique enables rapid learning of efficient social and domain-specific policies in complex and non-deterministic environments. Critically, we arguethat this evaluation demonstrates that SPARC is generic and can be successfully applied to a broad range of difficult human-robot interaction scenarios.

Relation to WP - This work contributes to Tasks T4.5-T4.6.

A.3 Bartlett et al. (2019) Recognizing Human Internal States: A Conceptor-Based Approach

Bibliography - Bartlett, M., Hernández García, D., Thill, S., and Belpaeme, T. (2019) Recognizing Human Internal States: A Conceptor-Based Approach. *Submitted to Social Robots in Therapy and Care workshop at the IEEE/ACM International Conference on Human-Robot Interaction 2019, Daegu, South Korea.*.

Abstract - The past few decades has seen increased interest in the application of social robots to interventions for Autism Spectrum Disorder as behavioural coaches [16]. We consider that robots embedded in therapies and interventions could also provide quantitative diagnostic information by observing patient behaviours. The social nature of ASD symptoms means that, to achieve this, robots need to be able to recognize the internal states their human interaction partners are experiencing, e.g. states of confusion, engagement etc. Approaching this problem can be broken down into two questions: (1) what information, accessible to robots, can be used to recognize internal states, and (2) how can a system classify internal states such that it allows for sufficiently detailed diagnostic information? In this paper we discuss these two questions in depth and propose a novel, conceptor-based classifier. We report the initial results of this system in a proof-of-concept study and outline plans for future work.

References

[1] James Kennedy, Séverin Lemaignan, Caroline Montassier, Pauline Lavalade, Bahar Irfan, Fotios Papadopoulos, Emmanuel Senft, and Tony Belpaeme. Child speech recognition in human-robot interaction: evaluations and recommendations. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 82–90. ACM, 2017.

[2] Alexandros Potamianos and Shrikanth Narayanan. Robust recognition of children's speech. *IEEE Transactions on speech and audio processing*, 11(6):603–616, 2003.

[3] Matteo Gerosa, Diego Giuliani, and Fabio Brugnara. Acoustic variability and automatic recognition of childrens speech. *Speech Communication*, 49(10-11):847–860, 2007.

[4] Romain Serizel and Diego Giuliani. Deep-neural network approaches for speech recognition with heterogeneous groups of speakers including children. *Natural Language Engineering*, 23(3):325–350, 2017.

[5] Paul Viola and Michael Jones. Robust real-time object detection. In *International Journal of Computer Vision*, 2001.

[6] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[7] B. Irfan, M. Garcia Ortiz, N. Lyubova, and T. Belpaeme. Multi-modal incremental bayesian network with online learning for open world user identification. 2019.

[8] Séverin Lemaignan, Charlotte ER Edmunds, Emmanuel Senft, and Tony Belpaeme. The pinsoro dataset: Supporting the data-driven study of child-child and child-robot social dynamics. *PloS one*, 13(10):e0205999, 2018.

[9] Haibin Cai, Bangli Liu, Zhaojie Ju, Serge Thill, Tony Belpaeme, Bram Vanderborght, and Honghai Liu. Accurate eye center localization via hierarchical adaptive convolution. In *British Machine Vision Conference*. British Machine Vision Association, 2018.

[10] Arvid Kappas. The science of emotion as a multidisciplinary research paradigm. *Behavioural processes*, 60(2):85–98, 2002.

[11] Aditi Ramachandran, Chien-Ming Huang, and Brian Scassellati. Give me a break!: Personalized timing strategies to promote learning in robot-child tutoring. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 146–155. ACM, 2017.

[12] M. Bartlett, D. Hernández García, S. Thill, , and T. Belpaeme. Recognizing human internal states: A conceptor-based approach. 2019. Submitted to Social Robots in Therapy and Care workshop at the IEEE/ACM International Conference on Human-Robot Interaction 2019, Daegu, South Korea.

[13] Herbert Jaeger. Conceptors: an easy introduction. *arXiv preprint arXiv:1406.2671*, 2014.

[14] MAJ Vlaar, Josje Verhagen, O Paz, and PPM Leseman. Comparing l2 word learning through a tablet or real objects: What benefits learning most? 2017.

[15] Florian Eyben, Martin Wöllmer, and Björn Schuller. Opensmile: the Munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1459–1462. ACM, 2010.

[16] Brian Scassellati, Henny Admoni, and Maja Matarić. Robots for use in autism research. *Annual review of biomedical engineering*, 14:275–294, 2012.

frontiers

Multi-modal Incremental Bayesian Network with Online Learning for Open World User Identification

Bahar Irfan [1,*], **Michaël Garcia Ortiz** [2], **Natalia Lyubova** [3] **and Tony Belpaeme** [4,1]

[1] *Centre for Robotics and Neural Systems, University of Plymouth, Plymouth, UK*
[2] *AI Lab, SoftBank Robotics Europe, Paris, France*
[3] *Prophesee, Paris, France*
[4] *IDLab - imec, Ghent University, Ghent, Belgium*

Correspondence*:
Bahar Irfan
bahar.irfan@plymouth.ac.uk

Word Count: 7532

ABSTRACT

Word Count: 187

User identification is an important step in creating a personalised long-term interaction with robots, such as in domestic applications, education, or rehabilitation. It requires learning the users continuously and incrementally, and possibly starting from a state without any known user. In this paper, we describe a multi-modal incremental Bayesian network with online learning to be applied in such scenarios. Face recognition is used as the primary biometric, and it is combined with ancillary information, such as gender, age, height and time of interaction obtained from proprietary algorithms, with a hybrid normalisation approach that combines the optimal normalisation method for each parameter to improve the recognition. We introduce the long-term recognition performance loss that weighs the importance of correct estimations of known users to the incorrect estimations of unknown users for optimising the parameters of the network. We generated multi-modal datasets with 200 users with random or periodic interaction times, that simulates an HRI scenario to evaluate our approach. The results show that the proposed network decreases the loss compared to face recognition alone and increases the identification rate substantially up to 40% in open-set and closed-set scenarios.

Keywords: Open world recognition; Bayesian network; soft biometrics; incremental learning; online learning; multi-modal dataset; long-term user recognition; Human-Robot Interaction

1 INTRODUCTION

User identification is an important step towards achieving and maintaining a personalised long-term interaction with robots. For instance, a user would need to be identified for providing personalised assistance in rehabilitation therapy (Lara et al., 2017). When a robot is first deployed it will start from a "tabula rasa", with no prior knowledge of users, and users will be encountered over a sometimes extended

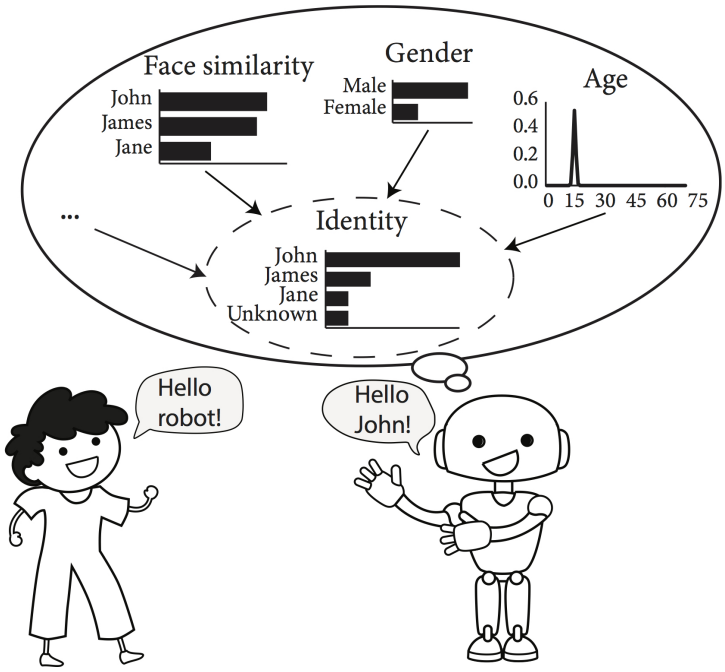

Figure 1. Robots can make use of multi-modal information to recognise users more accurately in long-term interactions

period of time. Hence, the system has to identify enrolled and "unknown" users, which is known as *open-set identification*. Open-set identification is a well-established (Scheirer et al., 2013; Jain et al., 2014; Scheirer et al., 2014), but in a real-world setting, these unknown users might need to be added into the system for future recognition. One solution is to re-train the entire system after introducing a novel user. However, this requires storing the previous samples, which could create a prohibitive computational burden in long-term deployments. Furthermore, it would require a significant amount of time to retrain with a growing number of users and samples (Bendale and Boult, 2015). Instead, the system should allow scaling and support the incremental learning of new classes, which is termed *open world recognition* (Bendale and Boult, 2015).

Face recognition (FR) has been the most prominent technique in biometric identification due to its non-intrusive character. Even though most state-of-the-art methods use deep learning based approaches (Taigman et al., 2014; Sun et al., 2014; Parkhi et al., 2015; Schroff et al., 2015), only a few approaches exist for open-set recognition (Bendale and Boult, 2016; Ge et al., 2017). Most models are not suitable for open world recognition due to the *catastrophic forgetting* problem, which refers to the drastic loss of performance on the previously learned classes when a new class is introduced (McClelland et al., 1995; McCloskey and Cohen, 1989; Parisi et al., 2018). The existing approaches that could help to overcome this problem often require a part of the previous data for re-training, which might not be available.

Incremental learning is not sufficient for adapting to the changes in the environment. For instance, an algorithm designed for open world recognition may not be able to recognise a person after a new haircut, because the model is not updated for known samples. Humans show a good model for recognition because they can continuously adapt to changing circumstances by updating their prior beliefs, known as *online learning*, and also use multi-modal information instead of a single biometric for estimation of an identity, such as recognising a person from the voice in a dark room. Most robots are also suitable for multi-modal recognition, as they have multiple sensors and perception algorithms (as shown in Fig. 1), which allow them to recognise users even when data are inaccurate or noisy, such as, a blurry image or illumination

changes (Wójcik et al., 2016). Moreover, the combination of multi-modal data can help to overcome issues related to similarities between users[1], by differentiating on additional available information, for example, age and gender. Such ancillary physical or behavioural characteristics, called *soft biometrics*, can be used to improve the recognition performance (Jain et al., 2011; Dantcheva et al., 2016). Combining multi-modal recognition with online learning can improve the recognition further in time. For instance, a user can be initially mistaken for another in certain circumstances, but these variations can be learned over time and combined with other modalities to improve recognition where FR fails.

In this paper, we extend our earlier work (Irfan et al., 2018) that proposed a multi-modal weighted Bayesian Network (BN) with online learning, combining soft biometrics (gender, age, height and time of interaction) with a primary biometric (face recognition) for open world user identification in human-robot interaction (HRI). Our main contribution is the extension of this method to take in multi-modal information, typically available in HRI, to markedly increase user identification and subsequently improve the user experience in long-term interactions for a large number of users. We make the following contributions (source code is available[2]):

- introducing *long-term recognition performance loss*
- formulating proposed online learning in terms of Expectation Maximization (EM) and Maximum Likelihood (ML)
- combining optimal normalisation methods for each parameter in the BN in a *hybrid* approach
- creating a multi-modal long-term user recognition dataset with 200 users of varying characteristics based on IMDB-WIKI dataset (Rothe et al., 2016) for evaluating the model with a large number of users

Obtaining a dataset which encapsulates a diverse set of characteristics for a large number of users over long-term interactions is a laborious task in HRI. However, it is important to optimise the parameters of the model on a dataset with a large number of users for applicability to various situations and domains of application. Thus, we created a simulated dataset of 200 users using images selected from the IMDB dataset containing 20k users, allowing us to compare two types of application scenarios: (1) patterned interaction times in a week modeled through a Gaussian mixture model, where the user will be encountered certain times in specific days, which applies to HRI in rehabilitation and education areas, and (2) random interaction times represented by a uniform distribution, such as in domestic applications with companion robots, where the user can be seen at any time of the day in the week. The experiments are conducted by using the proprietary algorithms of the Pepper robot[3] to obtain the multi-modal biometric information (face, gender and age), while the height and time of interaction are artificially generated to simulate a long-term HRI scenario.

The rest of the paper is organised as follows: Section 2 gives a brief overview of the current practice of open-world recognition, online learning, multi-modal biometrics algorithms, and user recognition in human-robot interaction (HRI). Section 3 describes the methodology and the structure of the proposed Bayesian network. Section 4 explains the procedure of the creation of the multi-modal long-term user recognition dataset. Section 5 presents the empirical evaluation of the proposed methods on the multi-modal closed-set and open-set datasets. Section 6 concludes with a summary of the work.

[1] https://www.wired.com/story/10-year-old-face-id-unlocks-mothers-iphone-x/

[2] Link will be available for the final version of this paper.

[3] https://www.softbankrobotics.com/corp/robots/

2 RELATED WORK

Our work lies at the intersection of open-world recognition, online learning, multi-modal biometrics, and HRI.

2.1 Open World Recognition

One of the first algorithms applied to open world recognition was the Nearest-Non Outlier (NNO) (Bendale and Boult, 2015), which modified Nearest Class Mean (NCM) (Mensink et al., 2013) for open-set classification and incremental learning. Another approach is the Extreme Value Machine (EVM) (Rudd et al., 2018) based on the Extreme Value Theory. However, both of these methods work with incrementally adding a batch of new classes (e.g. 100 at a time), as opposed to incremental learning of classes (one at a time). Similarly, the approach proposed in (Fei et al., 2016) is based on a center-based similarity space learning method and on the 1-vs-rest strategy of Support Vector Machines (SVM) for object classification.

2.2 Online Learning

Several Online Learning (OL) methods exist for various application areas (Gepperth and Hammer, 2016). In video-based recognition, Lee and Kriegman (2005) proposed an online learning algorithm of probabilistic appearances, but a generic prior model is necessary for this approach. De Rosa et al. (2016) used online learning in open world recognition for incremental learning of classification metric, the threshold for novelty detection and describing the space of classes. The approach was applied to three existing algorithms, namely, NCM, NNO and Nearest Ball Classifier (NBC) (Rosa et al., 2015). Their results showed that online learning increases classification performance.

2.3 Multi-Modal Biometrics

In a multi-modal biometric system, information from different identifiers, such as face recognition or gender identification, is fused via prior or post classification (Jain et al., 2005). Prior classification requires access to the features or the sensor values of the identifiers, which are generally not available for proprietary algorithms. For post classification, two approaches exist: *classification* and *combination* of confidence scores. Classification methods, such as neural networks and SVM, combine non-homogeneous data from individual classifiers into a feature vector for further classification without the need for preprocessing. In the combination approach, the individual matching scores from the identifiers are combined into a scalar score in three steps: (1) normalisation of scores into a common domain, (2) combination of scores based on Bayes decision rule and posterior probabilities, e.g. *sum* or *product rule*, and (3) thresholding for classification. The performance of these approaches depends on the method and the threshold chosen.

Bayesian approaches have been widely used for combining primary biometrics, such as face and speaker recognition (Bigün et al., 1997; Verlinde et al., 1999), as well as combining soft biometrics (Jain et al., 2004; Scheirer et al., 2011; Abreu and Fairhurst, 2011; Jain and Park, 2009; Zewail et al., 2004; Park and Jain, 2010). For instance, Jain et al. (2004) proposed a BN for combining fingerprints with soft biometric traits, namely, gender, ethnicity, and height. They used a fixed weighting scheme, where the biometrics with smaller variability and larger distinguishing capability were given more weight and achieved a slight improvement in recognition. Similarly, Scheirer et al. (2011) used a BN with Noisy-OR weighting that combines FR with ethnicity, hair colour and gender, and non-soft biometric contextual information, such as the occupation and the location of the person. Contrary to the work in (Jain et al., 2004) and our approach, they used the accuracy of the estimators to adjust the FR match score.

2.4 User Recognition in Human-Robot Interaction

Similar to biometric recognition, the most common approach for user recognition in HRI is through FR (Aryananda, 2001, 2009; Hanheide et al., 2008; Cruz et al., 2008; Gaisser et al., 2013). However, robots have a great advantage for multi-modal recognition due to the variety of multiple sensors that they can integrate. Soft biometrics are especially important because they allow non-intrusive recognition. However, only a few studies actually use soft biometrics. Martinson et al. (2013) used the weighted summation of soft biometrics (clothing, complexion and height) to identify the users within a short-term interaction from a group of only three users. Boucenna et al. (2016) gathered extensive data (100 images per person) during an imitation game and later evaluated the recognition offline using a Hebbian rule-based neural network. Ouellet et al. (2014) combined face recognition, speaker identification, and human metrology through Hampel estimators in closed-set identification using a substantial time for training (3.5 minutes) and a small number of participants (pretraining on 22, test on 7). Al-Qaderi and Rad (2018) combined face, body and speech information using a spiking neural network in the closed-set identification and have evaluated on a simulated dataset. These approaches do not apply for open-world recognition, hence, their methods are not easily comparable to ours.

Our previous work (Irfan et al., 2018) was the first approach in combining soft biometrics (gender, age, height and time of interaction) with a primary biometric (FR) to identify a user in real-time HRI. We introduced a multi-modal weighted BN that allows starting from a state of no known users to recognise unfamiliar users and incrementally learn them autonomously for open world recognition in HRI. Online learning was used for learning the likelihoods of the network from sequential data to improve the recognition over the long-term interactions. The weights of the network were optimised to minimise the number of incorrect recognitions. The *quality of estimation* measure was introduced to decrease the number of incorrect recognitions for unknown users. The results obtained in a user study with 14 participants over a four-week period showed a slight improvement in identification rate (up to 1.4% in open-set and 4.4% in closed-set recognition) compared to 90.3% of FR. The optimised weights suggested that age is the least effective soft biometric parameter, whereas height is the most effective one. Moreover, the BN performed worse with online learning. However, we concluded that the dataset might be biased towards the participants' characteristics due to the low number of participants and restricted age range, and an evaluation with a bigger dataset is necessary to fully understand the capabilities of the system.

This paper, thus, extends the work done in (Irfan et al., 2018), in evaluating the approach within a multi-modal long-term user recognition dataset, and optimising the weights of the BN through a *long-term recognition performance loss* criterion with a *hybrid* normalisation approach.

3 MULTI-MODAL INCREMENTAL BAYESIAN NETWORK

A Bayesian network is a probabilistic graphical model which represents conditional dependencies of a set of variables through a directed acyclic graph. BNs are suitable for combining scores of identifiers with uncertainties when the knowledge of the world is incomplete (Scheirer et al., 2011). The naive Bayes classifier model assumes conditional independence between the predictors, which is a reasonable assumption for a multi-modal biometric identifier as the individual identifiers do not affect each other's results.

We developed a weighted multi-modal incremental BN (MMIBN) (see Fig. 2) based on our previous work reported in (Irfan et al., 2018), integrating multi-modal biometric information for reliable recognition in open-world identification through a naive Bayes model. The primary biometric in our system is face

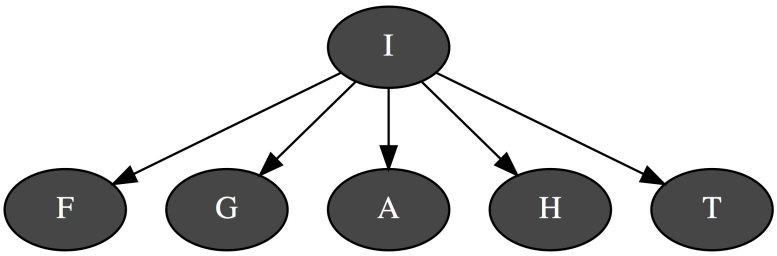

Figure 2. The naive Bayesian network model with identity (I), face (F), gender (G), age (A), height (H), and time of interaction (T) nodes.

recognition (F), which is fused with soft biometrics, namely, gender (G), age (A), and height (H) estimations and the time of interaction (T). We hypothesise that the integration of these soft biometrics will reduce the effects of noisy data as described in Section 1 and increase the identification rate. The pyAgrum (Gonzales et al., 2017) library is used for implementing the BN structure.

3.1 Structure

The number of states for each node depends on the modality: F and I nodes have n_e+1 states, where n_e is the number of enrolled (known) users. A and H nodes are restricted to the available range of the identifier, such as $[0, 75]$ for A and $[50, 240]$ for H; G has "female" and "male" states; T is defined by the day of the week and the time (the precise number of T states depends on the application as described later).

When a user is encountered, the corresponding multi-modal biometric evidence is collected from the identifiers. The FR provides similarity scores, which give the percentage of similarity of the user to the known faces in the database. Age, height, and time are assumed to be discrete random variables with a discretised and normalised normal distribution of probabilities, $N(\mu, \sigma^2)$, defined by (1), where V is the estimated value, Z is the standard score, and C is the confidence of the biometric indicator for the estimated value.

$$\mu = V, \quad P(\frac{-0.5}{\sigma} < Z < \frac{0.5}{\sigma}) = C \tag{1}$$

The time period (t_p) and its standard deviation (σ_t) can be set depending on the precision required in the application. For example, if the users in the application scenario will change every 5 minutes, then t_p = 5 min and σ_t = 15 min would be reasonable. On the other hand, in an HRI scenario, t_p = 30 min with σ_t = 60 min can allow better identification, because it is less likely to encounter users at the exact same time every day. Hence, we use the latter in this paper.

3.2 Weights of the Network

Soft biometric traits are characteristics that are not suited to uniquely identify an individual. We can assume that the population will have similar characteristics, but the distribution is unknown. However, some soft biometric features may contain more information about an individual than others, e.g. age is often more informative than gender. This can be modelled by using different weights for the parameters in a BN.

Weights (w_i) are used as the exponential to the likelihoods of the child nodes (X_i), similar to the work in (Zhou and Huang, 2006). The posterior probability $P(I^j|X_1,..,X_n)$ is approximated as shown in (2). I^j stands for the jth user ($I = j$), where I is the identity node.

$$P(I^j|X_1,..,X_n) \propto \frac{P(I^j)\prod_i P(X_i|I^j)^{w_i}}{P(X_1,..,X_n)} \tag{2}$$

As in (Jain et al., 2004), we assume that the identifiers perform equally well on all users. Therefore, the accuracy of an identifier is independent of the user, hence, equal priors are assumed for each of the identifiers. Therefore, the posterior probability simplifies to the equation shown in (3).

$$P(I^j|X_1,..,X_n) \propto P(I^j)\prod_i P(X_i|I^j)^{w_i} \tag{3}$$

Because the distribution of users over time is not known, one approach for determining $P(I^j)$ is to use adaptive priors using frequencies, as shown in (4), where n_{oj} is the number of times the user j is observed.

$$P(I^j) = \frac{n_{oj}}{\sum_j n_{oj}} \tag{4}$$

However, this can create a bias in the system towards the most frequently observed user as it affects directly the posterior probability, thus, may result in a decrease in the identification rate. Therefore, we assume that the probability of encountering the user j is equally likely as encountering the user m, hence, we assume equal priors for $P(I)$.

3.3 Quality of Estimation

Algorithms for open-set problems generally use a threshold (e.g. over the highest probability/score) to determine if the user is already enrolled or "unknown". However, the resulting posterior probabilities in a BN can be low due to the multiplication of the conditionally independent modalities and vary depending on the number of states. Hence, we use the two-step ad hoc mechanism introduced in (Irfan et al., 2018) to transform the BN to allow open-set recognition. (1) "Unknown" (U) state is used in both F and I nodes. The similarity score in FR of U is set to the FR threshold (θ_{FR}), such that when normalised, the scores below/above the threshold will have lower/higher probabilities than U. This allows to maintain the threshold for the FR system in use. (2) We use the confidence measure called the *quality of the estimation* (Q). Given the evidence y_t at time t, it compares the highest posterior probability (P_w) to the second highest (P_s), as shown in (5). The difference between the probabilities decreases, as the number of enrolled users (n_e) increases since $\sum_j P(I^j|y_t) = 1.0$. A similar method was used in (Filliat, 2007) for estimating the quality of localisation based on different images.

$$Q = [P_w(I^j|y_t) - P_s(I^j|y_t)] * n_e \tag{5}$$

If Q is less than the determined threshold (θ_Q), or U has the highest posterior probability, the identity is classified as unknown. Otherwise, the identity is estimated with a maximum a posteriori (MAP) estimation, given in (6).

$$j^* = \begin{cases} U, & \text{if } Q < \theta_Q \text{ or} \\ & P(I^U|y_t) > P(I^j|y_t) \text{ for all } j \\ \arg\max_j P(I^j|y_t), & \text{otherwise} \end{cases} \quad (6)$$

3.4 Incremental Learning

In HRI scenarios, it is desired to allow the user to enrol in the system, such that he/she can be recognised at the next encounter. For this, we use an online system, where a user is able to enrol by entering the name, gender, birth year, and height, and then a photo of the user is taken by the robot. This information is gathered to have the ground truth values for recognition, and for setting the initial likelihoods.

Initially, the system starts as "tabula rasa", where there are no known users. BN is formed when the first user is enrolled: one state for the new user and one for U for I node. The initial likelihood for F is set to be much higher for the true values as shown in (7), where w_F is the weight of the face variable, and n_e is the number of enrolled users. The value was found based on preliminary experiments.

$$P(F^k|I^j) = \begin{cases} 0.9^{w_F}, & \text{if } k = j \\ [0.1/(n_e - 1)]^{w_F}, & \text{otherwise} \end{cases} \quad (7)$$

The remaining likelihoods are set using the prior knowledge that the user entered in a similar structure to the evidence, i.e. for age, height and time variables with a discretised and normalised normal distribution, $N(\mu, \sigma^2)$, where μ is the true value (e.g. age of the person), and σ is the standard deviation of the identifier. Gender is set at $[0.99^{w_G}, 0.01^{w_G}]$ ratio, which is experimentally found. For the unknown state, $P(X_i^k|I^U)$ is set to be uniformly distributed as an unknown user can be of any age, height and be recognised at any time of the day, except for the face node, which follows (7).

When a new user is enrolled, the BN is expanded by adding a new state to the I and F nodes. $P(F^k|I^j)$ for each previous state in I (including U) is updated by appending the value corresponding to $k \neq j$ condition in (7), and then the probabilities are re-normalised. The likelihoods of G, A, H and T nodes for the previously enrolled users remain the same. This scalability feature removes the need to retrain the network when a new user is introduced, hence, the time complexity is decreased, which can be crucial if the new user is introduced at a later step (e.g. after 1000 users). More precisely, if each image corresponding to $\overline{n_o}$ average number of observations per user was to be recognised again after a new user is added to the face database, it would take a significant amount of time to expand the network compared to scaling, since $n_e * \overline{n_o} * \mathcal{O}(FR) \gg n_e * \mathcal{O}(1)$ updates, where $\mathcal{O}(FR)$ is the time complexity of the FR algorithm. In order to allow the network to make meaningful estimations, in the first few recognitions (here, we chose 5 recognitions), the identity is declared as unknown, regardless of the estimated identity.

3.5 Online Learning of Likelihoods

The BN parameters are generally determined by expert opinion or by learning from data (Koller and Friedman, 2009). The former can cause incorrect estimations if the set probabilities are not accurate enough. The latter, for which Maximum Likelihood (ML) estimation is commonly used, is not possible when the BN is constructed with incomplete data. One solution is to use offline batch learning, however, it requires storing data that can cause memory problems in long-term interactions. Another approach is to update the parameters as the data arrive, which is termed online learning. Variants of the Expectation Maximization

(EM) algorithm with a learning rate (EM(η)) (Bauer et al., 1997; Cohen et al., 2001; Lim and Cho, 2006; Liu and Liao, 2008) were proposed for online learning in BN.

We use a BN where the likelihoods are updated through EM(η) with an adaptive η (learning rate) based on ML estimation, similar to Voting EM (Cohen et al., 2001). Adopting the notation in (Bauer et al., 1997), where θ^t_{ijk} represents $P(X_i = x^k_i|I^j)$ at time t, the formulation is given in (8). The difference between voting EM and our approach is that we work with continuous probabilities due to uncertainties in the identifiers. We will refer to the proposed BN with online learning as MMIBN-OL.

$$\theta^{t+1}_{ijk} = \begin{cases} \eta_j P_{\theta^t}(x^k_i|y_t, I^j) + (1-\eta_j)\theta^t_{ijk}, & \text{if } P(I^j) = 1 \\ \theta^t_{ijk}, & \text{otherwise} \end{cases} \tag{8}$$

Combining ML estimate to achieve an adaptive learning rate (given in (9)) allows the learning rate to depend on the observation of the user j (n_{oj}), which is more reliable than using a fixed rate for all users. Also, each observation of the user creates a progressively smaller update on the likelihoods, such that, the effect of a new observation decreases as the number of recognitions of the user increases.

$$\eta_j = \frac{1}{n_{oj} + 1} \tag{9}$$

Supervised learning is necessary to achieve accurate online learning, i.e. the identity of the user should be known for updating the corresponding likelihoods, which can be achieved in HRI by asking for a confirmation of the estimated identity.

If the user j is not previously enrolled in the system, $P(F^k|I^U)$ is updated before updating $P(F^k|I^j)$ for each k. However, the likelihoods of gender, age, height, and time remain the same for U, to ensure uniform distribution.

3.6 Long-Term Recognition Performance Loss

Detection and Identification Rate (DIR) (fraction of correctly classified probes (samples) within the probes of the enrolled users ($\mathscr{P}_{\mathscr{E}}$), given in (10)) and False Alarm Rate (FAR) (fraction of incorrectly classified probes within the probes of the unknown users ($\mathscr{P}_{\mathscr{U}}$), given in (11)) are the standard metrics for open-set identification (Phillips et al., 2011).

$$DIR = \frac{|\{\arg\max_j P(I^j|y_t) = j | j, j \in \mathscr{P}_{\mathscr{E}}\}|}{|\mathscr{P}_{\mathscr{E}}|} \tag{10}$$

$$FAR = \frac{|\{\arg\max_j P(I^j|y_t) = j | k, j \in \mathscr{P}_{\mathscr{E}}, k \in \mathscr{P}_{\mathscr{U}}\}|}{|\mathscr{P}_{\mathscr{U}}|} \tag{11}$$

In other words, DIR represents the "true positive" (TP) of enrolled users, in which the current probe (referring to the multi-modal biometric sample) belongs to a user that is previously enrolled and identified correctly, within the fraction of probes belonging to the enrolled users. FAR serves as a "false positive" (FP) for unknown users, that is, the probe belongs to an unknown user, but he/she is identified as an enrolled user. However, TP and FP are notions of *verification* problems, in which the probe is compared against a claimed identity, thus, measures for F1-score or accuracy are generally not applicable to *open-set*

identification. Instead, the trade-off between DIR and FAR that depends on the threshold of the identifier, is generally represented by a Receiver Operating Characteristic (ROC) curve. The standard practice in biometric identification is to determine the desired FAR, which would then set the threshold and henceforth the DIR.

Depending on the biometric application, the cost of incorrectly identifying a user as known may be very different from the cost of incorrect identification of the enrolled user (Jain et al., 2011). For short-term interactions, in which a user will be encountered 1-2 times, FAR is as important or more important than DIR. However, for long-term interactions, the users will be encountered a greater number of times. Thus, correctly identifying a user (in a closed-set) becomes more important than correctly identifying that the user is unknown (open-set). Hence, we introduce the *long-term recognition performance loss* (L) that creates a balance between DIR and FAR based on the average number of observations per user ($\overline{n_o}$), as presented in (12), where α is the ratio of importance of DIR compared to FAR.

We optimise the weights of the BN through the loss function, for gender (w_G), age (w_A), height (w_H) and time (w_T) in $[0, 1]$ range, along with quality (Q) that can change within the $[0, 0.5]$ range. Ideally $L = 0$, where all the unknown users are identified as such (FAR $= 0.0$) and the known users are correctly identified (DIR $= 1.0$).

$$\begin{aligned} L &= \alpha * (1 - DIR) + (1 - \alpha) * FAR \\ \alpha &= 1 - \frac{1}{\overline{n_o}} \end{aligned} \tag{12}$$

3.7 Normalisation Methods

The scores from each modality must be normalised into a common range (e.g. $[0, 1]$) to ensure a meaningful combination. It is important to choose a method that is insensitive to outliers and provides a good estimate of the distribution (Jain et al., 2005), such as, minmax (MM), tanh (Hampel et al., 1986) (TH), softmax (Bishop, 2006) (SM), and norm-sum (dividing each value by the sum of values) (NS). We introduce *hybrid* normalisation (H) which combines the methods that achieve the lowest loss for each modality.

3.8 Extendability

The presented approach uses only one primary biometric, hence, in the absence of facial information, the image is discarded and the user is not recognized since soft biometric information would not be sufficient to estimate the identity. However, the system can be extended with other primary biometric traits, such as voice and fingerprint, and other soft biometrics, such as the location of interaction, eye colour and gait, to improve the recognition.

The proposed approach does not require heavy-computing, therefore, it is suitable for use on commercially available robots. We use this system on Pepper and NAO[4] robots for our experiments. These robots are operated by NAOqi[5] software, which includes different modules that allowed us to extract face similarity scores, gender, height and age estimations from a single image. However, the network is applicable to any identifier software on any platform. The estimations from these modalities are fed into the network. The internal states of the proprietary algorithm are inaccessible, hence, we assume that the gender and age

[4] https://www.softbankrobotics.com/corp/robots/

[5] http://doc.aldebaran.com/2-5

estimations are not used to obtain the face similarity scores, and they are conditionally independent from the FR results, even though they are obtained from the 2D image through the same module within NAOqi. Height is obtained from another module using the 3D camera in Pepper.

4 MULTI-MODAL LONG-TERM USER RECOGNITION DATASET

To the best of our knowledge, the only publicly available dataset that contains the soft biometrics used in our system (except for the time of interaction) with a dataset of faces is BioSoft (Sadhya et al., 2017). However, due to the low number of subjects (75), and the lack of numeric height values, we decided to create our own multi-modal long-term user recognition dataset.

For images, we use the IMDB-WIKI dataset (Rothe et al., 2016) which contains more than 500k images of celebrities with gender and age labels. We randomly sampled 200 celebrities out of 20k celebrities, choosing only celebrities which have more than 10 images each corresponding to the same age. The resulting dataset contains 101 females, 98 males and one transgender. In the dataset, each image of the user was chosen from the same year in order to simulate an open-world HRI scenario, where the users will be met in consecutive days or weeks. The images that correspond to an age that is within the five most common ages in the set were randomly rejected during the selection. The resulting age range is 10-63, with the mean age of 33.04 (SD 9.28). We assume that the IMDB dataset offers a diverse set of characteristics.

In the scope of this work, we use single-user recognition within the images, i.e. only one user is assumed to be present in each image. Hence, we use the cropped faces of the IMDB dataset, and we clean the dataset in three steps: we remove (1) images with a resolution lower than 150x150, (2) images without a face detected by NAOqi, (3) images that erroneously correspond to another person.

We assume that the average number of times a user will be observed is $\overline{n_o} \geq 10$, which is a reasonable assumption for long-term HRI. Hence, we create two datasets: (1) DT, where each user is observed exactly ten times, e.g. ten return visits to a robot therapist, and (2) DA, in which each user is encountered a different amount of times (10 to 41 times).

Pepper uses its 2D camera for face, gender and age recognition through NAOqi software, which can also work offline on images. In addition, the 3D camera is used for height estimation which requires a real person in front of the robot. Therefore, we had to artificially create height data, especially as (Irfan et al., 2018) found the height to be the most important soft biometric in determining the identity. To keep the data realistic, a Gaussian noise with σ = 6.3 cm found in (Irfan et al., 2018) was added to the true heights of the users obtained from the web.

Two types of distribution are considered for the time of interaction: uniform (U) for random interaction times and Gaussian mixture model (G) with three curves for users seen at certain times of the day in a week, resulting in a total of four datasets (DT_U, DT_G, DA_U, DA_G). The clean datasets of images and the resulting datasets are available online[6].

The initial datasets are divided into training and closed-set test using 100 users with 80-20% ratio of the data: 800 samples in DT and 2308 in DA for training, and 200 samples in DT, 620 in DA for closed-set. The open-set test is created from the remaining 100 users (800 samples in DT for equal comparison, 2280 in DA). The open-set evaluation is made by introducing these samples after the training dataset, i.e. the previous 100 users are enrolled in the system, and recognised multiple times before the introduction of the new users. However, the results for the open-set are evaluated only on the test set, not including the

[6] Link will be available for the final version of the paper.

training results. Weights (w_G, w_A, w_H, w_T) and quality of estimation (Q) are optimised using Bayesian optimisation[7] for 303 iterations over 5-fold cross-validation on the training set for each dataset and for each of the normalisation methods.

5 EVALUATION

In this section, we evaluate our proposed model based on the hypotheses presented in Section 5.1. Initially, the parameters of the multi-modal Bayesian network are optimised for open-world recognition in long-term interactions in Section 5.2. Using those parameters, the model is compared to face recognition and soft biometrics on the multi-modal long-term user recognition datasets for the training set, closed-set and open-set tests in Section 5.3.

5.1 Hypotheses

H1 Our proposed multi-modal BN will reduce the long-term recognition performance loss L and improve DIR compared to face recognition alone.

H2 Online learning will reduce L and improve DIR.

H3 Hybrid normalisation will outperform the individual normalisation methods.

H4 When the time of interaction is uniformly distributed (in DT_U and DA_U datasets), the loss will be higher.

H5 Optimised weight for the time in online learning will be low in uniformly distributed time datasets.

5.2 Optimisation of Parameters

In this section, we present our empirical evaluations to obtain the optimised parameters of our system on the described datasets: face recognition threshold (θ_{FR}), normalisation methods (Fig. 3), weights of the network and the quality of estimation (Fig. 4), and the performance loss according to FAR (Fig. 5).

5.2.1 Face Recognition Threshold

The loss parameter α in (12) should be set to find the optimum FR threshold (θ_{FR}) and optimise the parameters in our network. As α increases, the fraction of correct recognitions of enrolled users (DIR) increases, but the fraction of the incorrect recognitions of unknown users (FAR) will also increase. Based on our average number of observations assumption $\overline{n_o} = 10$ for long-term interaction, α becomes 0.9. For applications with fewer observations per user, α can be set accordingly.

If the highest similarity score is below the value of θ_{FR}, the identity is classified as unknown in FR. We examined how θ_{FR} influences the long-term recognition performance for the NAOqi FR, and noticed a decrease in performance for $\theta_{FR} > 0.4$. Hence, we chose $\theta_{FR} = 0.4$ for our network, such that, the similarity score of U will be as high as possible to decrease the FAR, in agreement with (Irfan et al., 2018).

5.2.2 Normalisation Methods

The long-term recognition performance loss of the normalisation methods in (Irfan et al., 2018), namely, minmax (MM), tanh (TH), norm-sum (NS), and softmax (SM), are compared with the introduced hybrid (H) normalisation (as shown in Fig. 3). Each normalisation method was evaluated using the datasets for each modality separately and the method that gave the lowest loss was used in the hybrid normalisation: norm-sum for face, gender, and height; tanh for age; softmax for time. The results in Fig. 3 show that hybrid

[7] https://thuijskens.github.io/2016/12/29/bayesian-optimisation/

normalisation provides the lowest loss in all datasets, followed by hybrid with online learning ($H - OL$), providing support for H3. H4 also holds true, as the loss is higher in nearly all of the normalisation methods when the time of interaction is uniformly distributed.

However, H2 is rejected for hybrid normalisation as the online learning increases the loss, as compared to the other normalisation methods for the Gaussian time datasets.

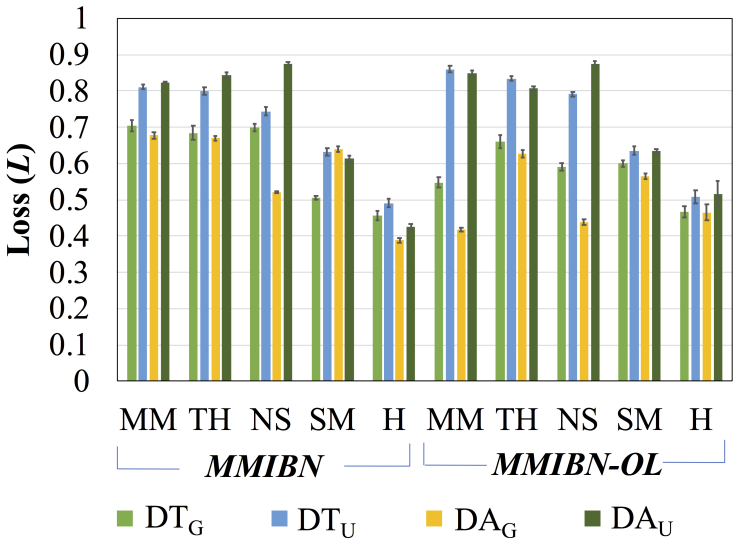

Figure 3. Comparison of loss in the training sets for normalisation methods with optimised weights: MM (minmax), TH (tanh), NS (norm-sum), SM (softmax), and H (hybrid). Lower loss is better. Standard deviation values of 5-fold cross-validation are shown with error bars.

5.2.3 Weights and Quality of Estimation

It seems to be self-evident that in the case of uniformly distributed time of interaction, online learning would provide worse results because the information provided by time will be unreliable. Hence, the optimisation should find a lower weight for the time parameter for MMIBN-OL (H5). The parameters corresponding to the optimum loss, presented in Fig. 4, show otherwise. w_T for the uniform time is higher than that of the Gaussian for online learning in both datasets.

In general, based on the relatively high weights, age seems to be the most important parameter, and height the least. This is in contrast with the findings in (Irfan et al., 2018). Since we are using a bigger dataset with varying characteristics, we can conclude that our results are more applicable to real-world deployments. However, it is important to note that the presented results are dependent on the defined loss function, the noise level of the identifiers and α. Hence, by adjusting α, setting a FAR, or using other algorithms for the identifiers, a different set of weights can be achieved with lower/higher FAR and consequently lower/higher DIR, as in Fig. 5 for DA_G.

The optimised quality of estimation (Q) was found to be less than 0.1 in each condition. The underlying reason is the disagreement of the modalities, which can decrease the differences in posterior probabilities because the results are combined through the product rule in the BN. When the modalities agree with high confidences (probabilities), the quality can be very high (e.g. see Figure 7 in Section 5.3) with Q = 7.44 for the probe of the second user.

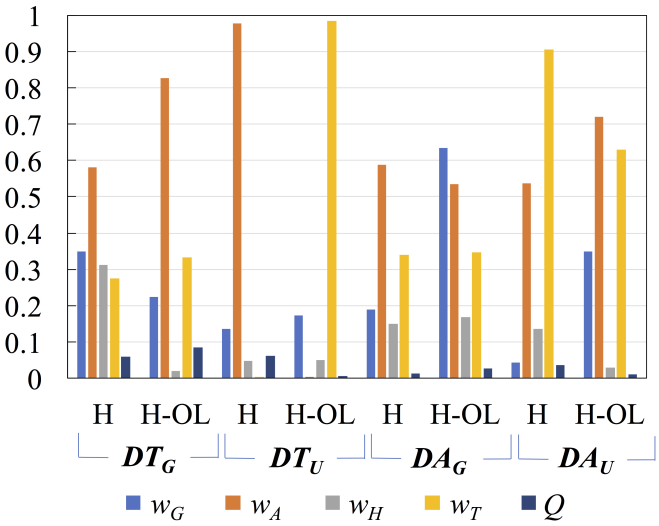

Figure 4. Optimised weights for gender (w_G), age (w_A), height (w_H), and time of interaction (w_T) and quality of estimation (Q) for hybrid normalisation for ten samples (DT) and all samples (DA) datasets with Gaussian ($_G$) and uniform ($_U$) times.

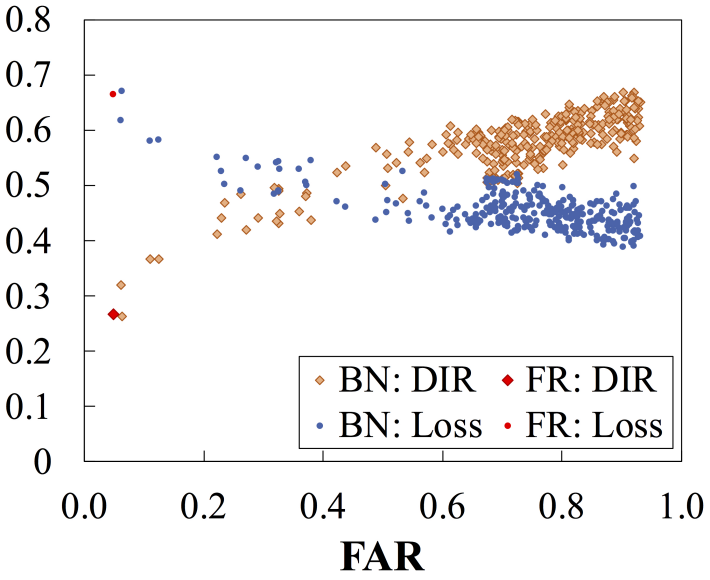

Figure 5. ROC curve for MMIBN-hybrid (BN) in the all samples dataset with Gaussian times (DA_G), with long-term recognition performance loss and DIR for varying FAR, for the Bayesian optimisation of weights and quality of estimation for 303 iterations over 5-fold cross-validation. Face recognition (FR) values are given for comparison. As DIR increases, loss decreases, but FAR increases. The loss parameter (α) can be adjusted or a FAR can be set to obtain a different set of weights.

5.3 Recognition Results

Lastly, we present the results of the training, closed-set and open-set datasets for our proposed BN, face recognition (FR) and soft biometrics (SB), in Fig. 6.

The results show that H1 is supported for all datasets, i.e. the proposed approach decreases the long-term recognition performance loss to a large extent compared to FR. The increase in DIR is drastic, from 22 to 40%, doubling the DIR of FR. This is a striking result, compared to the results in (Irfan et al., 2018) that

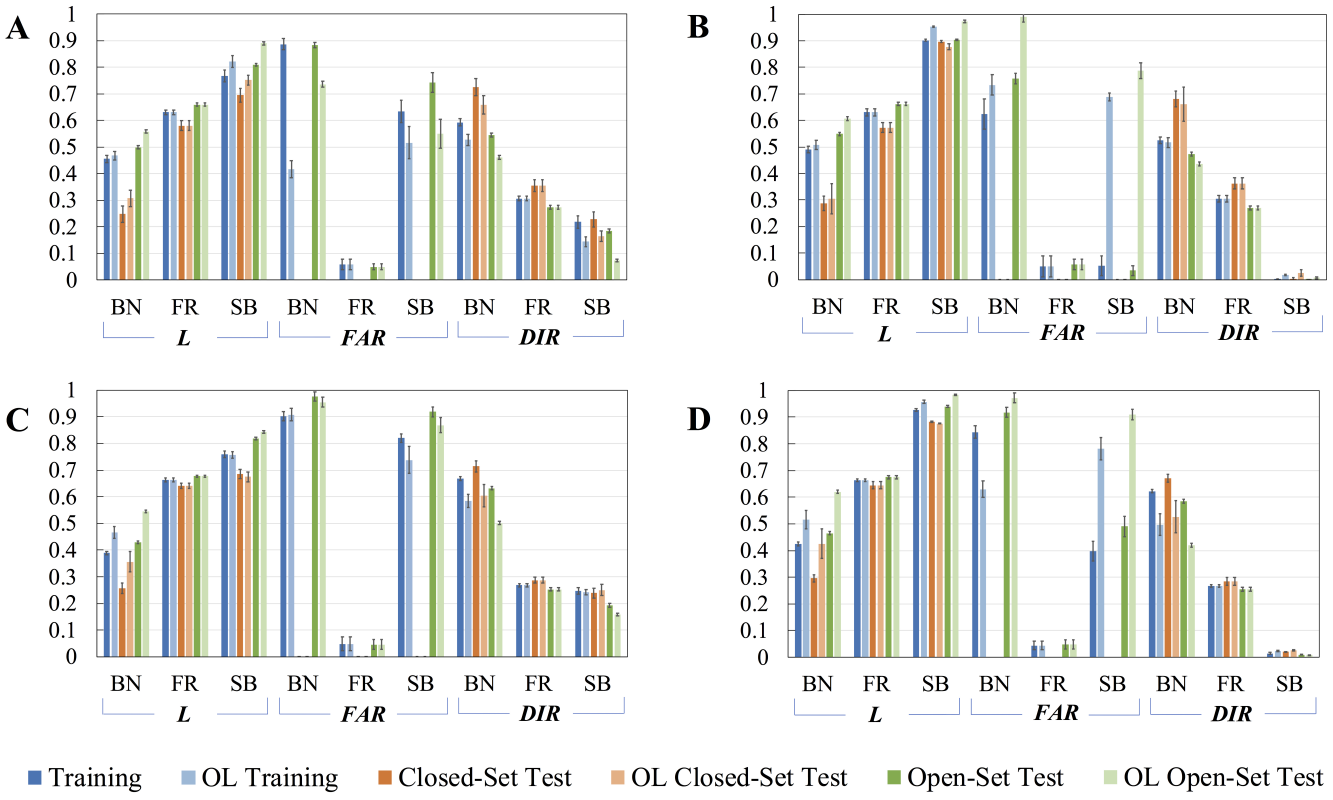

Figure 6. Comparison of loss (L), FAR and DIR for the proposed Bayesian network (BN), face recognition (FR), and soft biometrics (SB) on the presented datasets for training (100 users), closed-set test (100 users) and open-set tests (200 users) for ten samples (DT) and all samples (DA) datasets with Gaussian ($_{\mathrm{G}}$) and uniform ($_{\mathrm{U}}$) times: **(A)** DT_{G}, **(B)** DT_{U}, **(C)** DA_{G}, **(D)** DA_{U}. Standard deviation values of 5-fold cross-validation are shown with error bars.

showed an increase by only $1-4\%$. However, the higher number of users (100-200 users compared to 14) lowered face recognition DIR to a large extent from 90.3% to 30%, which could be one of the reasons of the substantial increase in DIR.

It should be noted that the increase in DIR provided by our network ($27-37\%$) is higher than the DIR of the soft biometrics ($19-22\%$). This shows that the soft biometric data are not sufficient to identify an individual, yet when combined with the primary biometric, they improve the identification rate considerably. This conclusion is supported by the datasets where the time of interaction is uniformly distributed. Due to the high variability of the time, the identification rate of SB is close to zero. However, in addition to FR, they improve the recognition by 22% in DT, and 35% in DA. This result along with the non-zero optimised weights support that the inclusion of age, gender and height modalities increases the recognition rates, suggesting that the visual modalities contain additional information to the FR, and confirming our initial assumption of conditional independence. Figure 7 shows examples from DA_{G} where the FR fails to recognise the user due to the low similarity score ($< \theta_{FR} = 0.4$), whereas, our proposed model identifies the user correctly based on the soft biometric information. The quality of estimation (Q) varies depending on the highest FR similarity score, as well as the disagreement between modalities. For example, for the third user (Sandra Oh), the highest FR similarity score (rank 1) is very low, corresponding to David Schwimmer who is 28 years old in the dataset, has a height of 185 with the enrollment time of interaction on Tuesday at 18:16. The age did not provide information to differentiate the user from the incorrect

estimation, whereas, the height and time of interaction increased the probability that the user is Sandra Oh, resulting in a correct estimation, but with a low quality score ($0.35 > \theta_Q = 0.013$). On the other hand, the second user (Gary Coleman) was identified correctly by FR with the highest similarity score close to, but slightly lower than θ_{FR}. This was enforced by the age estimation, and the time of interaction, which compensated for the incorrect recognitions of gender and height, to get a high quality score (7.44).

	Enrolment	*Probe:* 11		*Enrolment*	*Probe:* 2		*Enrolment*	*Probe:* 8	
									
	True Value	**Estimated Value**		**True Value**	**Estimated Value**		**True Value**	**Estimated Value**	
		FR	*BN*		*FR*	*BN*		*FR*	*BN*
ID	135	0 **[0.70]**	**135 [0.70]**	129	0 **[0.79]**	**129 [7.44]**	77	0 **[0.83]**	**77 [0.35]**
Name	Emilia Clarke	*FR (rank 1):* Angelina Jolie **[23.3%]**		Gary Coleman	*FR (rank 1):* Gary Coleman **[36.9%]**		Sandra Oh	*FR (rank 1):* David Schwimmer **[13.9%]**	
Gender	Female	Female **[72.7%]**		Male	Female **[88.1%]**		Female	Male **[66.3%]**	
Age	24	38 **[50%]**		10	7 **[100%]**		33	28 **[40%]**	
Height	157	152 **[8%]**		142	154.5 **[8%]**		168	172.7 **[8%]**	
Time	Saturday 10:15	Wednesday 17:35		Wednesday 13:41	Wednesday 13:53		Thursday 08:14	Thursday 07:57	

Figure 7. Examples of true values and estimated values of modalities from our multi-modal long-term user recognition dataset with Gaussian times (confidence values are given in brackets) using MMIBN hybrid. The highlights in red show the incorrect detection values. Face recognition (FR) was unable to recognise the users (0 represents unknown user) because the similarity scores were below the threshold value of 40%. Our proposed multi-modal Bayesian Network (BN) was successful (highlighted in green) in correctly identifying the users with varying quality of estimations (shown in brackets underneath the ID) as a result of the information gathered from the soft biometrics highlighted in blue. 8% confidence value of height corresponds to the σ = 6.3 cm in NAOqi. Images are taken from the IMDB-WIKI dataset (Rothe et al., 2016).

The open-set test results are comparable to the training set even though the number of users is dealing with has doubled in size, suggesting that the proposed approach and the optimised weights can generalise. The closed-set identification loss is considerably less than that of the training or open-set tests, due to the lack of unknown users (FAR = 0), and DIR is higher because the performance of the BN improves with time.

Face recognition alone, on larger datasets, typically has very poor recognition performance. In comparison, the proposed network has considerably higher FAR. However, as noted before, this is a result of the trade-off between recognition and spotting unknown people, which is visible in Fig. 5. Depending on the value of α in the loss function to ensure a higher number of correct recognitions in a long-term interaction.

Online learning provides comparable results for the DT datasets, but the loss is higher for DA. The underlying reason might be the accumulating noise in the identifiers. Therefore, for online learning, either identifiers with lower noise should be used or similar to the work (Cohen et al., 2001; Liu and Liao, 2008), the adaptive learning rate can decrease depending on the expectation and variance of the θ_{ijk}. Moreover, the confidence value of the identifiers or the quality of estimation can be used to determine if the likelihoods should be updated at each iteration, to avoid updating when the noise is high. The execution times per recognition (on a single CPU of Pepper robot) for online learning are considerably higher (mean (M) = 0.19 second, standard deviation (SD) = 0.0052) than the network without online learning (M = 0.04 s, SD = 0.0005). In addition, the average learned likelihoods (for 200 users) in online learning showed that the initial assumptions in (7) hold valid. The mean for face node was 0.913 (compared to the initial assumption of 0.9), with SD = 0.126. For the gender likelihood, M = 0.978 (the initial assumption was 0.99), SD = 0.058. Hence, it is sufficient and preferable to use the proposed approach without always updating the model whenever a user crosses the robot.

In general, it can be observed that FAR and DIR is higher in DA than in DT. The increase in DIR can be explained by the higher number of recognitions, which increases the performance over time. On the other hand, the increase in FAR can be due to the different optimised weights for each dataset (see Fig. 4). Since it is more likely that each user will appear a random number of times, we suggest to use the weights for the DA datasets; if the application is based on the users to come at specified times during a week (e.g. in a hospital), the optimised parameters for $\mathrm{DA_G}$ should be used, otherwise, it is better to use that of $\mathrm{DA_U}$ (e.g. for companion robots).

6 CONCLUSION

In this work, we wanted to use the different sensors a robot has to improve user identification and presented a multi-modal incremental Bayesian network with online learning and hybrid normalisation. We introduced a long-term recognition performance loss for optimising the parameters of the network that considers correctly identifying the enrolled users as more important than detecting unknown users. We have created a multi-modal dataset to simulate an HRI scenario, allowing us to evaluate the approach on a large number of users with varying characteristics. This dataset provides a proof of concept through which the parameters of the system can be optimised against. The results were generated by feeding simulated users to the Pepper robot's proprietary algorithms, thereby, providing real signals to our Bayesian Network.

The results show that the presented approach improves the identification rate substantially, and decreases the loss compared to face recognition alone. Furthermore, the introduced hybrid normalisation method outperforms the normalisation methods in (Irfan et al., 2018). However, contrary to our initial hypothesis, online learning decreases recognition performance and reduces the ability to spot unknown users, which could be due to the accumulating noise in the network. The optimised weights suggest that age is the most important identifier in soft biometric information, whereas, height is the least important. However, this result depends on the population characteristics and the performance of the face recognition and other artificial perception software.

It should be pointed out that the training set was used to optimise the parameters of the system, and the open-set evaluations showed that these parameters are valid and generalisable to other datasets. Closed-set evaluations showed that the system improves over time, however, the already high identification rates in the training sets prove that the proposed system works sufficiently well in open world recognition starting from a state of zero known users. The achieved low identification rates (70% in closed-set and 60% in open-set) arise from the notably low face recognition rates (30%), and thus, are not comparable to the results of the state-of-the-art closed-set face recognition approaches ($\sim 90\%$). However, this paper aims to show that the identification rate can be greatly improved by combining other available information (gender, age, height and time of interaction) without increasing the complexity of the system on a commercial robot with low-computational power. In addition, the closed-set approaches are not applicable to incremental learning, which is vital for HRI. Moreover, our model allows using better quality components (e.g. for face or age) to increase the recognition results, and it can be applied on any robot. Thus, it is suitable to be applied on robots in the wild and in the long-term HRI studies as an initial step towards personalising the interaction.

CONFLICT OF INTEREST STATEMENT

The research was conducted in University of Plymouth (UK) and SoftBank Robotics Europe (France) under the EU funding from APRIL, L2TOR and DREAM projects stated in detail in Section Funding. Author Michaël Garcia Ortiz is employed by the company SoftBank Robotics Europe and author Natalia Lyubova is employed by the company Prophesee (France). All other authors declare no competing interests.

AUTHOR CONTRIBUTIONS

BI, MO, NL and TB contributed to the conception and design of the study; BI wrote the algorithm described in this paper; BI created the dataset; BI performed the analysis of the data; BI prepared the figures; BI wrote the first draft of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This work has been supported by the EU H2020 Marie Skłodowska-Curie Actions Innovative Training Networks project APRIL (grant 674868), the EU H2020 L2TOR project (grant 688014), and the EU FP7 DREAM project (grant 611391).

ACKNOWLEDGMENTS

The authors would like to thank Valerio Biscione for his valuable suggestions in the design of the Bayesian network, Pierre-Henri Wuillemin for his substantial help with the pyAgrum library, and Hoang-Long Cao for his contribution of artwork in the human-robot interaction diagram (Fig. 1).

DATA AVAILABILITY STATEMENT

The images used in our dataset belong to IMDB-WIKI dataset (Rothe et al., 2016), which is composed of images collected from the Internet. This dataset states that it is made available for academic research purposes only, and the copyright of the images belongs to the original owners. Consequently, the raw data of our dataset supporting the conclusions of this manuscript will be made available online by the authors, without undue reservation, to any qualified researcher, for the final version of this paper.

ETHICS STATEMENT

The authors declare that the research was conducted in the absence of any human or animal subjects. The biometric data used for the study is based on the images from the IMDB-WIKI dataset (Rothe et al., 2016), which are collected from the Internet.

REFERENCES

Abreu, M. C. D. C. and Fairhurst, M. (2011). Enhancing identity prediction using a novel approach to combining hard- and soft-biometric information. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 41, 599–607. doi:10.1109/TSMCC.2010.2056920

Al-Qaderi, M. K. and Rad, A. B. (2018). A multi-modal person recognition system for social robots. *Applied Sciences* 8. doi:10.3390/app8030387

Aryananda, L. (2001). Online and unsupervised face recognition for humanoid robot: toward relationship with people. In *IEEE-RAS International Conference on Humanoid Robots*

Aryananda, L. (2009). Learning to recognize familiar faces in the real world. In *2009 IEEE International Conference on Robotics and Automation (ICRA)*. 1991–1996. doi:10.1109/ROBOT.2009.5152362

Bauer, E., Koller, D., and Singer, Y. (1997). Update rules for parameter estimation in Bayesian networks. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence* (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.), UAI'97, 3–13

Bendale, A. and Boult, T. (2015). Towards open world recognition. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1893–1902. doi:10.1109/CVPR.2015.7298799

Bendale, A. and Boult, T. E. (2016). Towards open set deep networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1563–1572. doi:10.1109/CVPR.2016.173

Bigün, E. S., Bigün, J., Duc, B., and Fischer, S. (1997). Expert conciliation for multi modal person authentication systems by Bayesian statistics. In *Audio- and Video-based Biometric Person Authentication. AVBPA 1997. Lecture Notes in Computer Science*, eds. J. Bigün, G. Chollet, and G. Borgefors (Springer), vol. 1206. doi:10.1007/BFb0016008

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer-Verlag New York, Inc.)

Boucenna, S., Cohen, D., Meltzoff, A. N., Gaussier, P., and Chetouani, M. (2016). Robots learn to recognize individuals from imitative encounters with people and avatars. *Scientific Reports* 6, 19908. doi:10.1038/srep19908

Cohen, I., Bronstein, A., and Cozman, F. G. (2001). *Online Learning of Bayesian Network Parameters*. Tech. Rep. HPL-2001-55 (R.1), HP Laboratories

Cruz, C., Sucar, L. E., and Morales, E. F. (2008). Real-time face recognition for human-robot interaction. In *2008 8th IEEE International Conference on Automatic Face Gesture Recognition*. 1–6. doi:10.1109/AFGR.2008.4813386

Dantcheva, A., Elia, P., and Ross, A. (2016). What else does your biometric data reveal? a survey on soft biometrics. *IEEE Transactions on Information Forensics and Security* 11, 441–467. doi:10.1109/TIFS.2015.2480381

De Rosa, R., Mensink, T., and Caputo, B. (2016). Online open world recognition. *CoRR* abs/1604.02275

Fei, G., Wang, S., and Liu, B. (2016). Learning cumulatively to become more knowledgeable. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM), KDD '16, 1565–1574. doi:10.1145/2939672.2939835

Filliat, D. (2007). A visual bag of words method for interactive qualitative localization and mapping. In *Proceedings 2007 IEEE International Conference on Robotics and Automation*. 3921–3926. doi:10.1109/ROBOT.2007.364080

Gaisser, F., Rudinac, M., Jonker, P. P., and Tax, D. (2013). Online face recognition and learning for cognitive robots. In *2013 16th International Conference on Advanced Robotics (ICAR)*. 1–9. doi:10.1109/ICAR.2013.6766498

Ge, Z., Demyanov, S., Chen, Z., and Garnavi, R. (2017). Generative openmax for multi-class open set classification. *CoRR* abs/1707.07418

Gepperth, A. and Hammer, B. (2016). Incremental learning algorithms and applications. In *European Symposium on Artificial Neural Networks (ESANN)*

Gonzales, C., Torti, L., and Wuillemin, P.-H. (2017). aGrUM: a Graphical Universal Model framework. In *International Conference on Industrial Engineering, Other Applications of Applied Intelligent Systems*. Proceedings of the 30th International Conference on Industrial Engineering, Other Applications of Applied Intelligent Systems. doi:10.1007/978-3-319-60045-1_20

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions* (Wiley)

Hanheide, M., Wrede, S., Lang, C., and Sagerer, G. (2008). Who am I talking with? A face memory for social robots. In *2008 IEEE International Conference on Robotics and Automation, ICRA 2008, May 19-23, 2008, Pasadena, California, USA*. 3660–3665. doi:10.1109/ROBOT.2008.4543772

Irfan, B., Lyubova, N., Garcia Ortiz, M., and Belpaeme, T. (2018). Multi-modal open-set person identification in HRI. In *2018 ACM/IEEE International Conference on Human-Robot Interaction Social Robots in the Wild workshop* (ACM)

Jain, A., Nandakumar, K., and Ross, A. (2005). Score normalization in multimodal biometric systems. *Pattern Recognition* 38, 2270–2285. doi:10.1016/j.patcog.2005.01.012

Jain, A. K., Dass, S. C., and Nandakumar, K. (2004). Soft biometric traits for personal recognition systems. In *International Conference on Biometric Authentication*. No. 3072 in LNCS, 731–738. doi:10.1007/978-3-540-25948-0_99

Jain, A. K. and Park, U. (2009). Facial marks: Soft biometric for face recognition. In *2009 16th IEEE International Conference on Image Processing (ICIP)*. 37–40. doi:10.1109/ICIP.2009.5413921

Jain, A. K., Ross, A. A., and Nandakumar, K. (2011). *Introduction to Biometrics* (Springer Publishing Company, Incorporated), chap. Multibiometrics. 209–258

Jain, L. P., Scheirer, W. J., and Boult, T. E. (2014). Multi-class open set recognition using probability of inclusion. In *Computer Vision – ECCV 2014*, eds. D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Springer International Publishing), 393–409

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques* (The MIT Press), chap. Parameter Estimation. 717–782

Lara, J. S., Casas, J., Aguirre, A., Munera, M., Rincon-Roncancio, M., Irfan, B., et al. (2017). Human-robot sensor interface for cardiac rehabilitation. In *2017 International Conference on Rehabilitation Robotics (ICORR)*. 1013–1018. doi:10.1109/ICORR.2017.8009382

Lee, K.-C. and Kriegman, D. (2005). Online learning of probabilistic appearance manifolds for video-based recognition and tracking. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. vol. 1, 852–859. doi:10.1109/CVPR.2005.260

Lim, S. and Cho, S.-B. (2006). Online learning of Bayesian network parameters with incomplete data. In *Computational Intelligence*, eds. D.-S. Huang, K. Li, and G. W. Irwin (Berlin, Heidelberg: Springer Berlin Heidelberg), 309–314

608 Liu, J. and Liao, Q. (2008). Online learning of Bayesian network parameters. In *2008 Fourth International*
609 *Conference on Natural Computation*. vol. 3, 267–271. doi:10.1109/ICNC.2008.651
610 Martinson, E., Lawson, W., and Trafton, G. (2013). Identifying people with soft-biometrics at fleet week.
611 In *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction* (IEEE
612 Press), HRI '13, 49–56
613 McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning
614 systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist
615 models of learning and memory. *Psychological Review* 102, 419–457
616 McCloskey, M. and Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The
617 sequential learning problem (Academic Press), vol. 24 of *Psychology of Learning and Motivation*. 109 –
618 165. doi:https://doi.org/10.1016/S0079-7421(08)60536-8
619 Mensink, T., Verbeek, J., Perronnin, F., and Csurka, G. (2013). Distance-based image classification:
620 Generalizing to new classes at near-zero cost. *IEEE Transactions on Pattern Analysis and Machine*
621 *Intelligence* 35, 2624–2637. doi:10.1109/TPAMI.2013.83
622 Ouellet, S., Grondin, F., Leconte, F., and Michaud, F. (2014). Multimodal biometric identification
623 system for mobile robots combining human metrology to face recognition and speaker identification. In
624 *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. 323–328.
625 doi:10.1109/ROMAN.2014.6926273
626 Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., and Wermter, S. (2018). Continual lifelong learning with
627 neural networks: A review. *CoRR* abs/1802.07569
628 Park, U. and Jain, A. K. (2010). Face matching and retrieval using soft biometrics. *IEEE Transactions on*
629 *Information Forensics and Security* 5, 406–415. doi:10.1109/TIFS.2010.2049842
630 Parkhi, O. M., Vedaldi, A., and Zisserman, A. (2015). Deep face recognition. In *British Machine Vision*
631 *Conference*
632 Phillips, P. J., Grother, P., and Micheals, R. (2011). Evaluation methods in face recognition. In *Handbook*
633 *of Face Recognition*, eds. S. Z. Li and A. K. Jain (Springer Publishing Company, Incorporated). 2nd
634 edn., 553–556. doi:10.1007/978-0-85729-932-1_21
635 Rosa, R. D., Orabona, F., and Cesa-Bianchi, N. (2015). The abacoc algorithm: A novel approach for
636 nonparametric classification of data streams. In *2015 IEEE International Conference on Data Mining*.
637 733–738. doi:10.1109/ICDM.2015.43
638 Rothe, R., Timofte, R., and Van Gool, L. (2016). Deep expectation of real and apparent age from a single
639 image without facial landmarks. *International Journal of Computer Vision (IJCV)* 126. doi:10.1007/
640 s11263-016-0940-3
641 Rudd, E. M., Jain, L. P., Scheirer, W. J., and Boult, T. E. (2018). The extreme value machine. *IEEE*
642 *transactions on pattern analysis and machine intelligence* 40, 762–768
643 Sadhya, D., Pahariya, P., Yadav, R., Rastogi, A., Kumar, A., Sharma, L., et al. (2017). Biosoft - a
644 multimodal biometric database incorporating soft traits. In *2017 IEEE International Conference on*
645 *Identity, Security and Behavior Analysis (ISBA)*. 1–6. doi:10.1109/ISBA.2017.7947693
646 Scheirer, W. J., de Rezende Rocha, A., Sapkota, A., and Boult, T. E. (2013). Toward open set recognition.
647 *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 1757–1772. doi:10.1109/TPAMI.
648 2012.256
649 Scheirer, W. J., Jain, L. P., and Boult, T. E. (2014). Probability models for open set recognition. *IEEE*
650 *Transactions on Pattern Analysis and Machine Intelligence* 36, 2317–2324. doi:10.1109/TPAMI.2014.
651 2321392

652 Scheirer, W. J., Kumar, N., Ricanek, K., Belhumeur, P. N., and Boult, T. E. (2011). Fusing with context:
653 A Bayesian approach to combining descriptive attributes. In *2011 International Joint Conference on*
654 *Biometrics (IJCB)*. 1–8. doi:10.1109/IJCB.2011.6117490
655 Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition
656 and clustering. *CoRR* abs/1503.03832
657 Sun, Y., Wang, X., and Tang, X. (2014). Deep learning face representation from predicting 10,000 classes.
658 In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 1891–1898. doi:10.1109/CVPR.
659 2014.244
660 Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). Deepface: Closing the gap to human-level
661 performance in face verification. In *Proceedings of the 2014 IEEE Conference on Computer Vision*
662 *and Pattern Recognition* (Washington, DC, USA: IEEE Computer Society), CVPR '14, 1701–1708.
663 doi:10.1109/CVPR.2014.220
664 Verlinde, P., Druyts, P., Cholet, G., and Acheroy., M. (1999). Applying Bayes based classifiers for decision
665 fusion in a multi-modal identity verification system. In *Intl. Symposium. on Pattern Recognition*
666 Wójcik, W., Gromaszek, K., and Junisbekov, M. (2016). Face recognition: Issues, methods and alternative
667 applications. In *Face Recognition - Semisupervised Classification, Subspace Projection and Evaluation*
668 *Methods*, ed. S. Ramakrishnan (InTech), chap. 02. doi:10.5772/61471
669 Zewail, R., Elsafi, A., Saeb, M., and Hamdy, N. (2004). Soft and hard biometrics fusion for improved
670 identity verification. In *Circuits and Systems, 2004. MWSCAS '04. The 2004 47th Midwest Symposium*
671 *on*. vol. 1, I–225. doi:10.1109/MWSCAS.2004.1353967
672 Zhou, Y. and Huang, T. S. (2006). Weighted Bayesian network for visual tracking. In *18th International*
673 *Conference on Pattern Recognition (ICPR'06)*. vol. 1, 523–526. doi:10.1109/ICPR.2006.1188

Teaching robots social autonomy from in-situ human guidance

Emmanuel Senft[1*], Séverin Lemaignan[2], Paul Baxter[3], Madeleine Bartlett[1], Tony Belpaeme[1,4]

[1]Centre for Robotics and Neural Systems, University of Plymouth, UK,
[2]Bristol Robotics Lab, University of the West of England, Bristol, UK,
[3]L-CAS, University of Lincoln, UK
[4]ID Lab - imec, University of Ghent, Belgium

[*]Corresponding author; E-mail: emmanuel.senft@plymouth.ac.uk

Striking the right balance between human control and robot autonomy is a core challenge in social robotics, both in technical and ethical terms. On the one hand, extended robot autonomy offers the potential for increased human productivity and the off-loading of physical and cognitive tasks; on the other hand making the most of human technical and social expertise, as well as maintaining accountability, is highly desirable. This issue is particularly sensitive in domains such as medical therapy and education where social robots hold substantial promise, but where there is a high cost to poorly performing autonomous systems, compounded with sensitive ethical concerns. We present an ecologically valid study evaluating SPARC, a novel approach addressing this challenge whereby a robot progressively learns appropriate autonomous behaviour from in-situ human demonstrations and guidance. Using online machine learning techniques, we demonstrate that the robot can effectively

acquire legible and congruent social policies in a high-dimensional child tutoring situation, from a limited number of demonstrations. By exploiting human expertise, our technique enables rapid learning of efficient social and domain-specific policies in complex and non-deterministic environments. Critically, we argue that this evaluation demonstrates that SPARC is generic and can be successfully applied to a broad range of difficult human-robot interaction scenarios.

Introduction

intro part 1: 'learn autonomy instead of program autonomy' – example of predictive typing[SL]

intro part 2: a very difficult special case: education[SL]

Tony to add a description of the education domain and how robots need autonomy, and how this approach might be a good way forward[TB]

In sensitive domains where social robots are expected to play a key role, such as education and therapy, the question of empowering the human user by allowing them to supervise and retain fully transparent control over the robots has to be constantly balanced with the contradictory expectation of an advanced level of robot autonomy. Additionally, the growing expectation is that robots should behave autonomously not only at a technical, task-specific level, but also at in terms of social interactions.

In this article, we look at one specific, yet difficult, instance of this problem: how domain experts (called hereafter human *teachers*) can transfer both technical and social skills to enable robots to successfully and autonomously interact with children in an educational task. The expectation is that a robot could gradually learn an adequate social behaviour by observing the human teacher, and would become increasingly autonomous in both task-level skills and social interactions. As the teacher starts to trust the robot's behaviour, she or he would pro-

gressively shift their workload to the robot. In such a scenario, the robot's technical and social policies would be co-constructed by the teacher during the learning phase, and the resulting (autonomous) robot behaviour would thus remain essentially transparent, thus predictable and trustworthy to the human teacher (*1*). Educational social robotics is a prototypical application domain in this regard: to be an effective support, the robot needs to exhibit satisfactory technical (didactic, i.e., subject knowledge) and social (pedagogic behaviour) skills, all while preserving the ability for a school teacher to oversee and, if needed, override the robot's behaviour.

justification for the *learning* of behaviours?[ES]

Learning Autonomy Instead of Programming Autonomy Learning social policies for interactions with humans brings specific requirements, not usually considered in machine learning:

R1 The robot has to exhibit, at all times, an acceptable (socially and physically safe) – if not perfectly appropriate – social and task-related behaviour. This starting from the onset of the learning/interaction.

R2 The robot needs to learn quickly, as gathering data points with humans is a slow and costly process.

R3 To be effective in real world scenarios (where the human experts teaching the robot are not the roboticists), the learning process must be practical, integrate well with the natural human routines and require limited engineering expertise.

Classically, two main methods exist for teaching robots, Reinforcement Learning (RL) (*2*) and Learning from Demonstrations (*3, 4*). One of the core mechanisms of RL is the combination of exploration and learning from errors. To be effective, this requires both exploration and error recovery to be cheap, thus RL approaches typically rely on simulators to quickly train the agent. Simulation is, however, often not an option for human-robot interaction, as no simulator

to date is able to reproduce at meaningful levels the complexity and unpredictability of human behaviours ref?[ES]. This effectively means that the robots have to be trained in the real world by interacting with humans. Exploring and recovering from errors in the real world, however, is expensive, and sometimes not possible at all. Not being able to fully recover from errors in HRI is the norm rather than the exception: when one observes that human-robot interactions almost always requires a level of trust, it becomes clear that if the human loses their trust in the robot due to a poor behaviour choice, resuming the interaction to its fullest might prove impossible. Such failure modes are called *catastrophic failures*, and limit the general applicability of classical RL to HRI (as this violates (R1)). Additionally, learning with RL is often a slow process requiring millions of step, thus violating (R2).

To palliate these limitations, robots can also learn from humans, which ensures that the robot's policy is appropriate to the current application during the learning process. *Learning from Demonstration* (*3,4*) is one classical approach that enables humans to teach skills to robots. However, it typically looks at kinaesthetic demonstrations in deterministic environments, where the human teacher usually relinquishes control and supervision of the robot once the physical skill is deemed to have been acquired by the robot. *Learning by Demonstration* is commonly found in the context of manufacturing, industrial robotics or cobotics. It has also been applied, in a few instances, to the learning of social, interactive behaviours (*5–7*). These approaches might lead to positive results in an autonomous phase, however, significant engineering work was required after gathering the demonstrations to transform them into a policy a robot could apply to interact autonomously. This additional step implies that technical experts are required to be present throughout the learning process to interpret demonstration data and create learning algorithms adapted to each environment. This constant requirement of experts violates (R3), thus limiting the usability of such an approach by members of the general population. In the ideal case, the robot would learn from the supervision online, and would be ready to interact

autonomously immediately after, without this additional engineering step.

One way to combine advantages from both online learning and supervised learning is Interactive Machine Learning (IML) (*8, 9*). IML involves the end-user in the learning loop and has the agent learn an appropriate behaviour online through a series of small improvements. For example, people can provide rewards on the robot's actions, similarly to classic RL (*10*). Involving the user in the learning process allows them to provide additional information to the learner, making use of human teaching expertise to speed up the learning and allow the robot to learn behaviours specially adapted to the user's desires. Additionally, keeping the user in the loop and actively involved in the teaching process allows the user to create a mental model of the robot, increasing the transparency of the robot behaviour and the trust the user has in the machine (*11, 12*). However, teachers could also be given more control over the robot by dynamically providing demonstrations, correction or additional information to the algorithm to speed up the learning further (*13, 14*). Furthermore, with this human control, the teacher can correct errors made by the algorithm before they impact the real world, providing safety to the learning process. However, while holding promise, there are very few demonstrations of IML applied to learning for social interactions with humans, robots learning online in HRI generally only adapt within predefined boundaries (*15, 16*). IML, and Interactive reinforcement learning in particular, has had limited success so far, and mostly in simple, low-dimensional and deterministic interaction domains (*11, 17*).

Our approach (called SPARC: Supervised Progressively Autonomous Robot Competencies (*18*)), unlike the methods discussed thus far, addresses the three requirements stated previously by defining a new interaction paradigm: the robot interacts directly with the environment under the supervision of a human teacher who has total control over the robot's behaviour. Initially, the robot is fully teleoperated by the human teacher in a Wizard-of-Oz fashion: the teacher can select actions for the robot to execute. The robot learns a policy from these expert

operations, and immediately starts to suggest actions to the teacher based on the policy it has learned up to that point. The teacher can confirm or override the robot's suggestions, and this feedback is fed to the learning algorithm to progressively refine the policy. In order to reduce the teacher's workload, actions proposed by the robot and not cancelled by the teacher are assumed to be acceptable, and executed after a short delay. This mechanism aims to limit the requirement for human intervention. The teacher only has to demonstrate actions and prevent incorrect actions from being executed. Thus, as the robot's behaviour improves, the robot proposes correct actions more often, reducing the need for demonstrations and corrections, and thereby the amount of input required from the teacher to achieve an efficient behaviour.

When applied to HRI, for example in the context of education, this translates into transforming a dyadic interaction {human teacher; learning child} into a triadic interaction {human teacher; robot; child}, where the teacher teaches the robot how to support the child's learning on-the-go, such that the robot can autonomously make appropriate use of a combination of didactic and pedagogic actions (Figure 1).

This approach shares similarities with predictive texting: the predictive texting engine makes right away suggestions based on a database of common words, and progressively adapts to its user, learning from their decisions, and proposing more appropriate actions. It is a case of a machine learning to better help the user, while keeping him or her in full control of the exact output. Similarly, SPARC allows humans to teach robots an interaction policy while keeping them in control but with a gradually diminishing workload until reaching a point where the agent is trusted enough to interact autonomously.

The conceptual simplicity of the paradigm makes it widely applicable to a range of social human-robot interactions beyond the specific educational scenario that we use as support in this article and to other fields such as classic machine learning for robotics or other application.

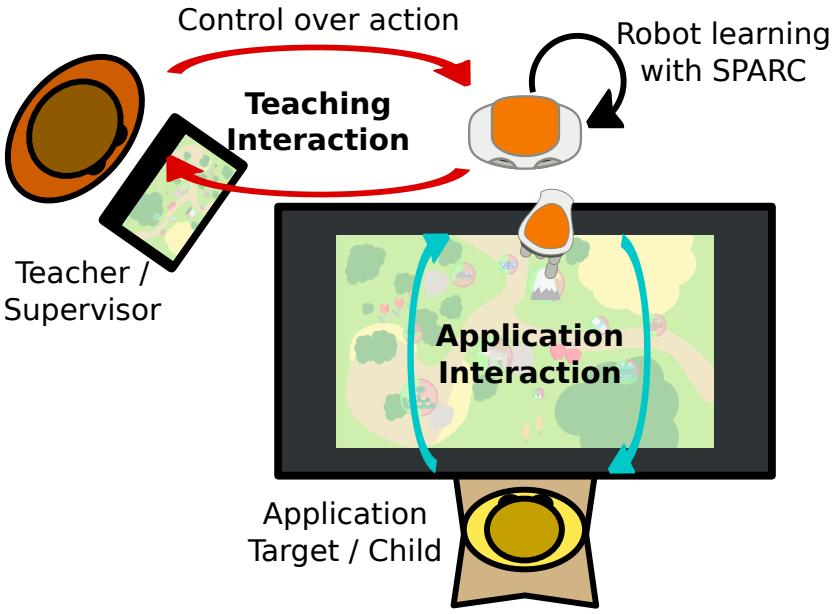

Figure 1: Diagram of the application of SPARC to HRI: a human teacher supervises a robot learning to interact with a second human partner, the target of the application. In the context of education, this application target would be a child learning new concepts.

Case study: Robots as Tutors for Children Social robots in educations have been explored in the last decade. Due to increased diversity in the classroom and budget constraints, teacher are no longer able to give personalized attention to pupils. One solution is to use a robot that takes the role of a tutor or teacher and offers personalized lesson or tutoring sessions. Recent studies have shown that social robots quite often are more effective than alternative technologies, such as tutoring software presented on a tablet or computer. The physical presence of the robots together with its social appearance promotes behaviours in the learner, such as increased attention and compliance, which are conducive to learning (*19*). However, robots for learning are often programmed with a limited repertoire of behaviors and they do not adapt to the specific learner or learning environment. Having a robot which can be operated initially by the teacher but then gradually takes over control, would offer a tutoring experience which is better tailored to the particular learner or learning task.

Study Introduction This paper's contribution is a study evaluating SPARC in a high-dimensional social task where 8 to 10 year old children learned about food webs through playing a game (Figure 2). In this game, 10 animals can be moved around in a touchscreen-based game environment; animals have energy and have to eat plant or other animals to stay alive and the child has to learn to balance animals diets to keep the ecosystem viable as long as possible. The role of the robot tutor is to guide the child using advice (such as keeping track of the animals' energy or indicating what animals eat) and social prompts (e.g. encouraging the child). The game logic and the tutoring interaction are jointly modelled as an optimisation problem with 210 continuous input values (last actions, distances between animals, etc.) and 655 potential output actions (motions, gestures, verbal encouragements, etc.).

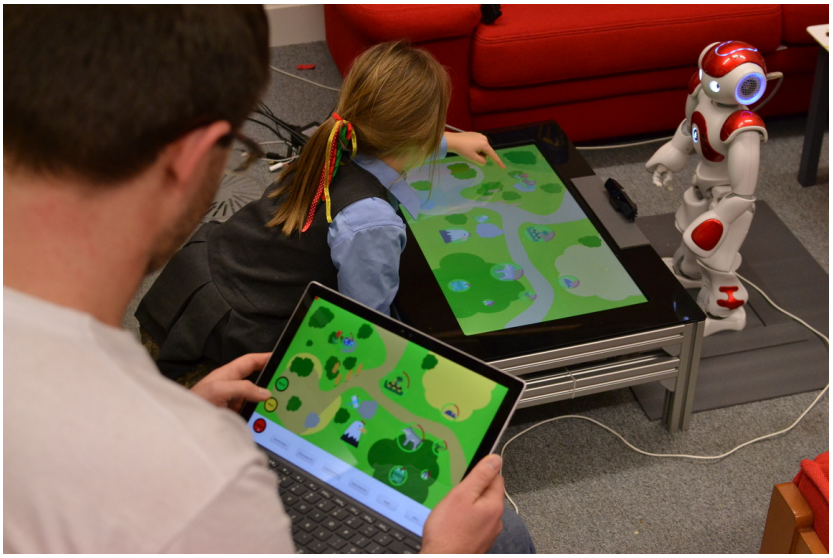

Figure 2: The setup used in the study: a child interacts with the robot tutor, with a large touchscreen sitting between them, displaying the learning activity; a human teacher provides guidance to the robot through a tablet and monitors the robots learning.

The interaction consists on four consecutive and independent game rounds, and knowledge tests before the first game, between the second and the third and after the fourth.

Our protocol includes three conditions, designed to assess the impact of applying the pro-

posed approach (SPARC) to this task. The control condition (*Passive condition*) uses a passive robot that only provides initial instructions and guidelines, but does not offer support during the learning game. The second, the *Supervised condition*, involves a robot which gradually learns from human demonstration how to provide support during the game by using SPARC. The third, the *Autonomous condition*, uses an autonomous robot which executes the policy learnt in the supervised condition, but without ongoing supervision. A range of metrics are used to evaluate performance in each condition, including the children's learning gain (using pre- and post-tests), within-interaction behaviour (e.g. robot's selected actions or child's behaviours) and the behaviour of the human teacher in interacting with the robot in a supervisor capacity.

Hypotheses Four hypotheses have been explored in this research:

H1 In the supervised condition, the teacher will be able to ensure an appropriate robot behaviour whilst teaching. We predict that the teacher will be able to use the interface to have the robot executing only desired actions in the supervised condition, thus leading to improvements in the children's behaviours.

H2 The autonomous robot will be able to interact efficiently during the game while exhibiting a social behaviour, and maintain the child's engagement during the learning task. We predict that the autonomous robot will reach a tutoring performance similar to the supervised robot and better than the passive robot.

H3 An active robot (supervised or autonomous) supports child learning: the learning gain in the passive condition will be inferior to the learning gain in the autonomous condition, which will be inferior to the learning gain in the supervised condition.

H4 Using SPARC, the supervisor's workload decreases over time: the number of corrected actions and the number of actions selected by the teacher decrease with practice, while

the number of accepted proposed actions increases.

H3 is motivated by the idea that the humans possess knowledge which should help the child to learn more from the game. By learning this knowledge, the autonomous robot should be able to partially replicate this effect, but without being able to match it, due to the limits of the algorithm, the interface and the limited time spent teaching.

Results

Example of a Session Table 1 presents an example of the first minute of a round. Propositions from the robot are in blue, and actions from the teacher in orange. For example, at t=16.9, the teacher accepted the proposition from the robot. Alternatively, in some cases, such as the suggestion at t=20.6, the teacher did not evaluate the action proposed by the robot, but simply selected another action. In that case, the action proposed is not evaluated and only the selected action is executed and used for learning. In other cases, such as at t=6.6, the algorithm suggested an action, the teacher decided to refuse it (by 'waiting'), before selecting it again after a short delay. Finally, at t=44.4 seconds, the teacher selected the action to move the mouse closer to the wheat, and after the robot moved the mouse, the child tried other animals and then fed the mouse with the wheat, this demonstrates how the actions from the robot could help the children to discover new connections between animals. This partially supports H1 ('In the supervised condition, the teacher will be able to ensure an appropriate robot behaviour whilst teaching') as we can see the teacher using the robot's propositions and selecting actions she deemed appropriate to the current situation.

Policy Comparison Figure 3 presents the number of actions of each type executed by the teacher (in the supervised condition) and by the autonomous robot. Both policies presented similarities: the action 'Move away' was almost never used, 'Move to' was never used, and

the supportive feedback ('Congratulation' and 'Encouragement') were used more often than 'Remind rules' or 'Drawing attention'. However, some dissimilarities were also present, for instance, the autonomous robot used more encouragements than congratulations while the teacher did the opposite. The autonomous robot also reminded the child of the rules more often and used the 'Move close' action less than the supervisor. These differences of actions are probably linked to the type of machine learning used; with instance-based learning, some data points will be used in the action selection much more often than others, which might explain these biases. Nevertheless, these results show that the autonomous robot based its actions on the one used in the supervised condition. The robot managed to learn a social and technical policy presenting similarities with the one displayed by the teacher, providing partial support for H2. The THRI paper already presented and discussed this part of the results, so it reduces the novelty of that part[ES]

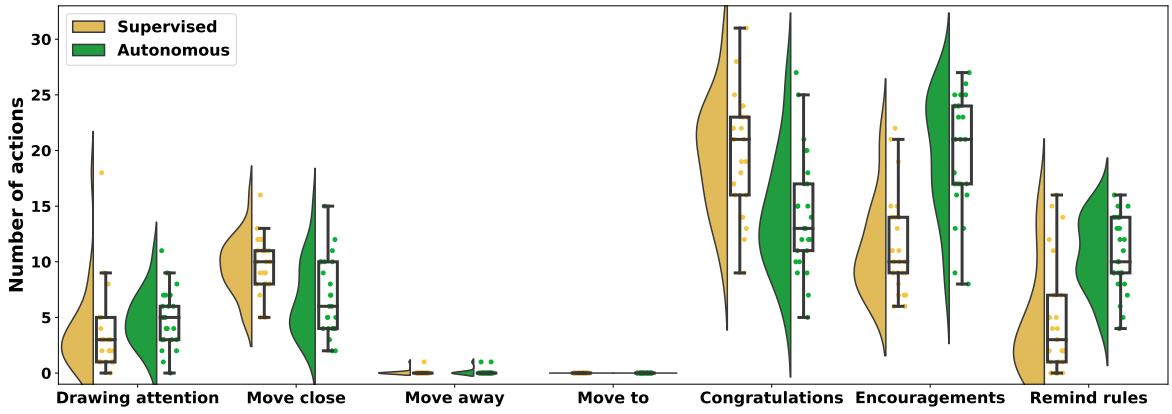

Figure 3: Comparison of the number of actions of each type executed by the robot in the autonomous and supervised conditions. While demonstrating differences, these two distributions show that the autonomous robot informed its choices of actions only from the demonstrations of the teacher.

Learning gains On average, children learned in all conditions, reaching a learning gain of 13% (passive: M=0.12 (SD=0.14), supervised: M=0.11 (SD=0.13), autonomous M=0.14 (SD=0.12)).

However, the robot's behaviour during the game did not have a meaningful impact on the children's learning gain (Bayesian ANOVA: $F(2, 72) = 0.337, \eta^2 = 0.01, p = .72, B_{10} = 0.15$) failing to support H3.

Game Metrics Multiple game metrics have been collected in the rounds of the game played by the children and can inform us on the effect of the robot's behaviour on the children during the game sessions.

Figure 4 and Table S1 show the evolution of the total number of different learning items encountered by the children across the four game rounds (corresponding to the number of different eating interactions created by the children). A Bayesian mixed-ANOVA showed an impact of the repetition (i.e. progress in the rounds of the game) and the condition on the number of different eating interactions produced by the children in the game (Bayesian mixed-ANOVA: repetition: $F(3, 216) = 6.75, \eta^2 = 0.08, p < .001, B_{10} = 78.8$, condition: $F(2, 72) = 5.2, \eta^2 = .13, p < .01, B_{10} = 5.7$). With additional rounds of the games, the children connected successfully more animal together. Post-hoc tests showed no significant difference between the supervised and the autonomous conditions (Bayesian Repeated-Measure ANOVA: $B_{10} = 0.154$), whilst differences were observed between the supervised and the passive conditions ($B_{10} = 512$) and between the autonomous and the passive conditions ($B_{10} = 246$). This indicates that, compared to the passive robot, the supervised robot provided additional knowledge to the children during the game, allowing them to create more useful interactions between animals and their food, receiving more information from the game, thus potentially helping them to get knowledge about what animals eat. Importantly, the autonomous robot managed to recreate this effect without the presence of a human in the action selection loop.

Figure 5 and Table S2 show the evolution of game duration across the four game rounds.

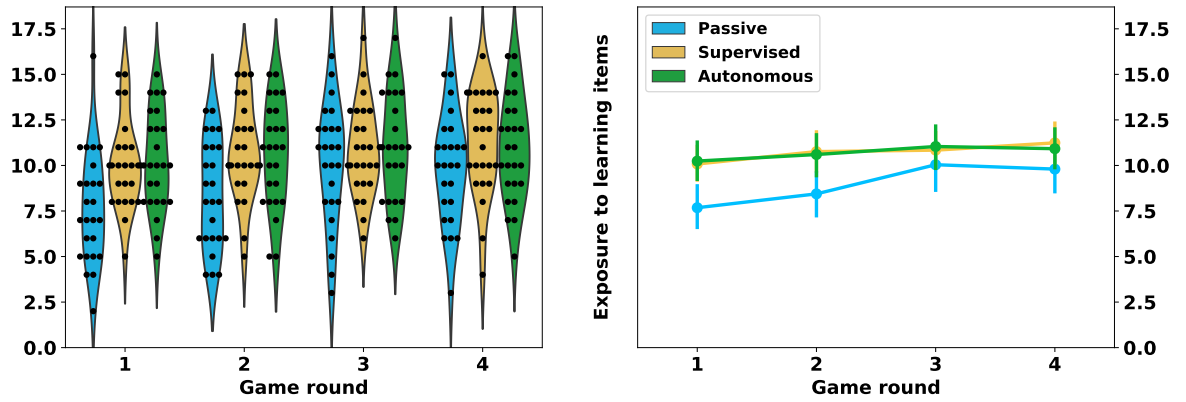

Figure 4: Number of different eating interactions produced by the children (corresponding to the exposure to learning items) for the four rounds of the game for the three conditions.

A Bayesian mixed-ANOVA showed inconclusive results on the impact of condition on game duration (Bayesian mixed-ANOVA: $F(2, 72) = 2.6, \eta^2 = 0.07, p = 0.08, B_{10} = 0.82$). Post-hoc tests showed no significant difference between the supervised and autonomous conditions (Bayesian Repeated-Measure ANOVA: $B_{10} = 0.287$), while differences were observed between the supervised and passive conditions ($B_{10} = 118$) and a trend towards a difference between the autonomous and passive conditions ($B_{10} = 2.9$). This indicates that children were better at the game in the supervised condition whereby animals were alive longer than in the passive condition. The autonomous robot learned and applied a policy tending to replicate this effect and without exhibiting differences with the supervised one.

However, the analysis showed no effect of the repetitions on game duration (Bayesian mixed-ANOVA with Huynh-Feldt correction: $F(2.4, 174.9) = 0.31, \eta^2 = 0.004, p = .82, B_{10} = 0.022$); the children did not manage to keep the animals alive longer with more practice at the game. One of the reasons was a partial ceiling effect at 2.25 minutes (see the red line on Figure 5). When not fed, animals would run out of energy in 2.25 minutes, so if children did not manage to feed at least 7 of the animals at least once before that time, the game would stop. As this might prove difficult to identify and achieve, many children did not manage to cross this

limit.

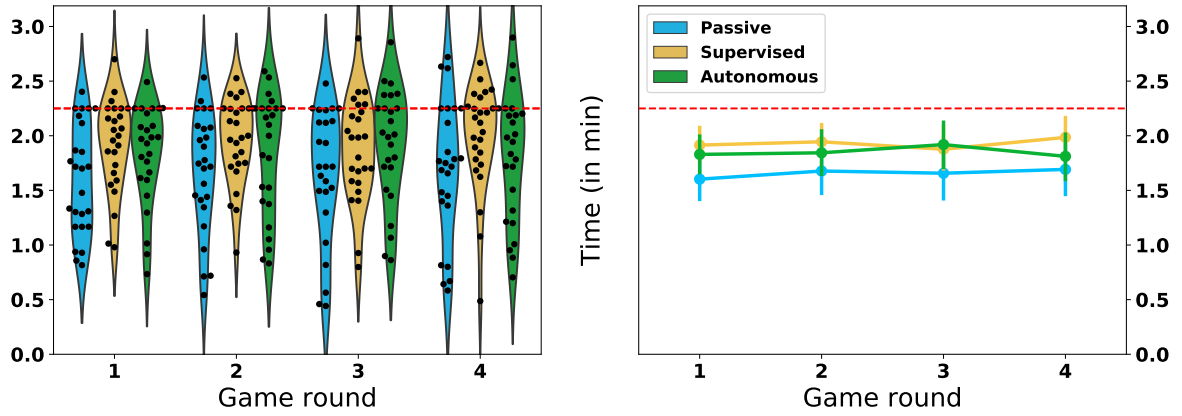

Figure 5: Interaction time for the four rounds of the game for the three conditions. The dashed red line represents 2.25 minutes, the time at which unfed animals died without intervention, leading to an end of the game if the child did not feed animals enough.

These game metrics suggest that the policy executed by the autonomous robot allowed children to achieve results in the game similar to those achieved when interacting with the supervised robot, and better results than with the passive robot. This provides support for H1 because, when teaching the robot to interact, the teacher managed to have a positive impact on the child from the onset. These results also support H2 because the autonomous robot tended to replicate this effect and provided an advantage to the children, unlike the passive robot.

Teaching the Robot Figure 6 presents the teacher's reactions to the robot's suggestions across all the supervised interactions. Contrary to our expectations, the number of accepted and refused suggestions, as well as teacher selections, stayed roughly constant throughout the interactions with the children. As the teacher aspect was a case study with a low number of data points and high variation between children, inferential statistical analysis, such as regression, would not be appropriate. On average, among the 4 rounds of an interaction, the teacher accepted 17.2 (SD=4.0) actions proposed by the robot (which represented 29.2% of the evaluated actions) and

41.7 (SD=11.1) were refused by the teacher per interaction. The teacher manually selected 25.8 (SD=5.8) actions per interaction. We would have expected these results to be different: with the learning, the number of accepted propositions should have increased and both the number of refused propositions and teacher selections should have decreased. It should, however, be noted that, due to this aspect being a case study, these results might not be replicated with another teacher. Should we add something about the teacher being naive about the hypothese, and just behaving as she preferred?[ES]

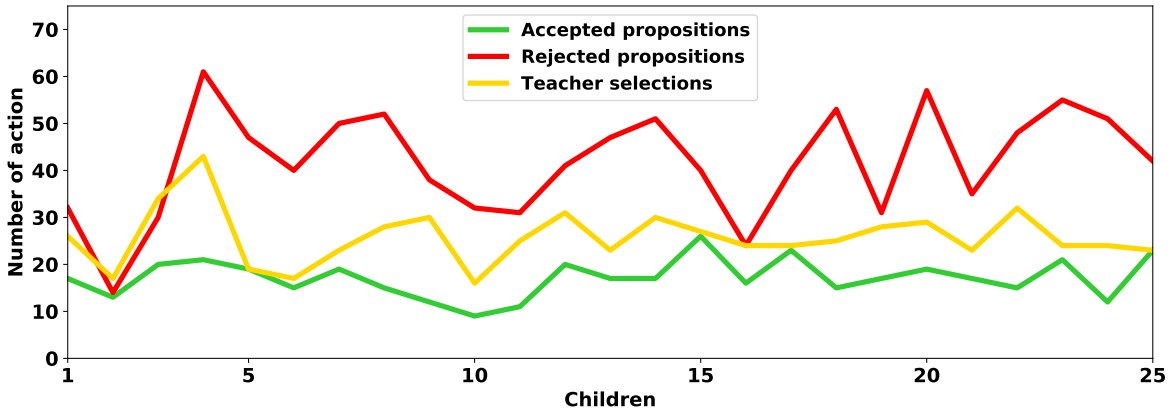

Figure 6: Summary of the action selection process in the supervised condition. The 'teacher selection' label represents each time the teacher manually selected an action not proposed by the robot.

In post-hoc discussion, the teacher reported three phases in her teaching (session numbers are indicative, the boundaries were not clear):

- First phase (sessions 1 to 5): she was not paying much attention to the suggestions, mostly focusing on having the robot executing a correct policy.
- Second phase (sessions 6 to 16): she was paying more attention to the suggestions but without giving them much credit.
- Third phase (sessions 17 to 25): she started to trust the robot more but without ever

trusting it totally.

More specifically, in a written diary she completed throughout the study, the teacher noted that she was "playing safe" in the very beginning, choosing to reject even valid robot suggestions, and sometimes finding the robot "overwhelming". The teacher's report suggested that this was due to the initially steep learning curve associated with supervising the robot; she felt she needed to get used to the supervision set-up before she could pay attention to the robot's suggestions. Interestingly, while the teacher's reports indicate that she felt she was trusting the robot more and allowing it to perform more of its suggested actions, this was not reflected in the actual frequency of accepted robot propositions.

The teacher did report a decrease of workload as she she progressed in the sessions number. This was supported by behaviours such as typing her observations on a laptop, while gazing at the interface in multiple interactions (especially at the start of a round). However, this decrease of workload seemed to be due mostly to the teacher getting used to the interaction, and not to the online learning and the improvement of the suggested propositions, invalidating H4.

Discussion

This study demonstrated that the robot successfully learned a behaviour providing support to children in the educational activity. This learning happened online, using as teacher a psychology student with no knowledge about the algorithm implementation. We showed that in little over three hours, a human could teach a robot a behaviour leading to 10 to 30% improvement in children's performance in the activity (number of learning items and length of interaction). While not showing a difference of learning gain or a reduction of workload over time. This study demonstrates that the principles behind SPARC allow an efficient teaching of social autonomy that can be done in the real world, at a human timescale and while maintaining an appropriate robot behaviour throughout the teaching and beyond, when the robot interacts autonomously.

We discuss hereafter the three main facets of our methodology: *in situ* learning; learning in *real-world environments*; and learning *social interactions*.

Learning in situ Similarly to other Interactive Machine Learning methods, SPARC learns online. This has associated advantages compared to offline learning such as classical Learning from Demonstration (LfD) or supervised learning.

First, online learning increases the usability by non-experts in computing: a robot learning online would only require one technical step, before the deployment. Once the robot is deployed and learning, it can be taught by anyone and then when an appropriate behaviour is reached, the robot can be deployed easily to interact autonomously. On the other hand, LfD methods applied to social HRI require significant engineering work before deploying the robot, but also once the demonstrations have been gathered to design a policy from them (*6, 7*). This additional step requiring technical expertise implies that these robots cannot be fully usable by non- experts in ML.

Second, learning in situ offers more flexibility in the range of possible applications. Given a general enough representation of the state and action spaces, an efficient algorithm and a clear interface, a single setup could result in a large spread of final behaviours without requiring a technical expert to step in at any time. For example, in this study, the same learning algorithm and world representations could be used to teach the robot a behaviour for typical children and a different behaviour for children with special needs. Thus online learning can allow each user to start with the same robot and controller, and define a policy suiting their specific needs. This online learning could also be used to refine seamlessly an already correct but imperfect policy.

Learning in situ allows end-users to design their own robotic controller without the requirement of any technical expertise. This might reduce the needs of engineering robots, thus making the process of designing a policy easier and more adaptive, potentially helping to democratise

the use of robots.

Learning in real-world & sensitive environments While the advantages of learning in situ apply to any IML methods, most of them provide only limited control to the teacher over the behaviour executed by the robot. This lack of control cannot ensure that the robot's behaviour will be safe for the interaction partners, the robot itself or its environment, thus reducing the applicability of such methods (*17*). However, by ensuring that the teacher vets each action executed by the robot, SPARC increases the range of application of IML to real sensitive environments. Furthermore, by having control over the behaviour, the teacher can speed up the learning, allowing robot to learn a useful behaviour in complex and stochastic environments in a reasonable amount of time.

In this study, the robot learned an efficient policy, comparable to the teacher's one in an ecologically valid (complex, under-specified, stochastic real-world interaction) and sensitive (education) environment. The environment had an input space of 210 dimensions and output action space of 655 actions. The interaction happened in the wild, in the school children are going to and the children displayed a number of unexpected behaviour the robot had to adapt to (such as intentional waiting, hectic play style...). The environment was also sensitive, incorrect hints or feedback from the robot could have caused distress or annoyance for the children. Thus, this learning situation was more challenging than many others where IML has been evaluated (often deterministic environment, with limited risks for failures... (*10, 11*)) or classical adaptive scenario for educational HRI (*15, 20*). But despite this, SPARC was successful both in the teaching phase (ensuring that the robot's behaviour was safe and useful from the onset) and in the autonomous phase (by demonstrating a behaviour comparable the teacher's policy and which had similar impacts on children).

Learning to be social By using robots learning from their users as they interact while keeping the users in control of the robot behaviour, SPARC has the potential to ease the use of learning robot, making them available to more situations and empowering a larger range of users. One of the specific situation addressed by this study by demonstrating its application to social environments.

Behaving socially is a challenge, there is no explicit set of rule defining a perfect behaviour, the interactants can behave in ways hard to anticipate and the recovery from errors can be costly if not impossible. SPARC has been specially designed to tackle these challenges: by learning online, the robot can progressively acquire a policy suited to the real interaction and this continuous supervision allows to cover cases not anticipated. Additionally the control provided to the teacher allows to maintain a social robot behaviour even in the early stages of the learning.

As demonstrated in this study, SPARC leads to promising results when applied to such a situation. When behaving in the autonomous condition, the robot allowed children to be better at the game (lasting longer) and to encounter more learning items compared to a passive robot. These results are similar to the ones observed when the teacher was controlling the robot. This indicates that the robot manage to re-enact a social policy similar to the one the teacher used, thus paving the way to robots learning to interact socially with humans.

Outlook Our results show that SPARC allows users non-experts in ML to teach a robot a social behaviour in situ, while interacting with children. This interaction framework is suitable for sensitive complex environment, allowing a safe teaching and the learning of an efficient robot behaviour.

Although our results demonstrated the opportunities provided by SPARC and other IML methods providing teachers with control, some limitations remains and motivate future work. This study did not show a decrease of the teacher's workload overtime. One of the main reason

was that the robot proposed actions too often, overloading the teacher and preventing her to take time to correctly evaluate each suggestion. Future work could explore ways to provide the teacher with more control not only on the overt robot behaviour (the one displayed in the application) but also in the teaching interaction (such as being able to control the time until autoexecution, waiting time between propositions or some meta parameter of the learning algorithm). Additionally, the children did not show difference of learning between the conditions. This can be explained by the fact that this task (game and test) was not validated before so there was no guarantee that learning gains from the test exactly represent the difference of knowledge accumulated by the children. This motivates the need of benchmark task for HRI that researcher can use to evaluate AI systems and more generally robot controllers. Further work could also explore way for the robot to go beyond the demonstrations, using the teacher to have a safe baseline but also fine tuning the behaviour with autonomous learning to reach super-human capabilities.

Conclusion This paper demonstrated the potential for SPARC to enable robots to learn from humans. This capability is especially useful in HRI as often, the knowledge of the robot behaviour lies in the hand of domain experts such as therapists or teacher. The classical way to design robotic controller requires multiple rounds of discussion between the engineers coding the behaviour and the domain expert. Robot learning from end-users (e.g. by using SPARC) would bypass these costly loops, allowing end-users to directly specify in an intuitive way an efficient controller adapted to their specific needs. Additionally, the control provided to the teachers makes this teaching safer, easier to use and more pleasant for the users.

The implications of this study are two-fold: first, we have demonstrated that, with an appropriate methodology, interactive machine learning can be successfully applied to transfer human expertise to an autonomous robot, in a short period of time, and in a high-dimensional and eco-

logically valid task. Second, for the first time, we have shown that not only domain-specific technical expertise, but also elements of social behaviours can be taught that way. not sure this is supported by the data??[ES]

Those two results are significant. The dynamic and stochastic nature of social interactions makes learning appropriate and contingent social behaviours a challenge for which classical machine learning approaches are ill-suited. We have shown here a path forward, and our approach makes it possible for autonomous social behaviours to be learnt in an online manner, gradually taking over the social interaction from the human operator.

Because the process fundamentally relies on having the human in the loop, it also holds considerable potential for sensitive applications of social robots, such as in e-health, assistive robotics or education.

Materials and Methods

Objective and Design We designed a study to test if SPARC could be used to teach a robot to interact in a complex, non-deterministic and real environment. In this study, a NAO robot [1] guided a child through a gamified tutoring session where the child had to interact with animals on a touchscreen to learn about food-webs. This study compared three conditions where the robot could be either *passive* (not providing any feedback or information to the child during the game), *supervised* (an adult, the teacher, was teaching the robot how to the support the child during the game) or *autonomous* (the robot interacted without supervision and executed autonomously the policy learned in the supervised condition).

Apparatus This study is based on the Sandtray paradigm (*21*): a child interacts with a robot via a large touchscreen located between them. By interacting with the touchscreen and the

[1] https://www.ald.softbankrobotics.com/en/robots/nao

robot, the child is expected to gain knowledge or improve some skills. Additionally, a teacher can control and teach the robot in the 'supervised' condition using a tablet. This results in a triadic interaction: a human, the teacher, knows how the robot should behave, can control it to execute an efficient behaviour and teach it how to interact with another human *in situ* by using SPARC (as shown in Figure 2).

Participants Children from five classrooms across two different primary schools in Plymouth (UK) were recruited to take part in the study. As both schools had an identical OFSTED evaluation (indicating that they provide similar educational environments), all the children were combined into a single pool of participants. Full permission to take part in the study and be recorded on video was acquired for all the participants via informed consent from parents. In total, 75 children were included in the final analysis, with 25 participants per condition (N=75; age: M=9.4, SD=0.72; 37 Female).

In the supervised condition, the robot's teacher was a psychology PhD student from the University of Plymouth, with limited knowledge of machine learning but with an understanding of human cognition. This teacher is now part of the authors, but at the time of the study the authorship was not considered and she was not involved in the study design. Consequently, while being knowledgeable about the protocol, she was unaware of the hypotheses tested and the implementation and had no incentive to bias the results to fit them. The teacher was instructed on how to control the robot using a Graphical User Interface on the tablet and the effects of each button. She experimented controlling the robot in two interactions (not included in the results analysis) to get used to the interface and controlling the robot. After these interactions, the algorithm was reset and the teacher started to supervise the robot for the supervised condition. No information about the learning algorithm or the representation of the state and no feedback about the optimal way of interacting or on her policy was provided before or during the study.

As such, this study involved, as teacher, a naive user not expert in ML and more similar to the general population of expected robot users than an expert in computing.

Protocol At the start of the interaction, a child was first introduced to the robot and told that they would play a game about food chains with the robot (cf. Figures S3.a). Then, they completed a quick demographic questionnaire and a first pre-test to evaluate their baseline knowledge (cf. Figures S3.b-e). After this test, and before starting the teaching game, the child had to complete a tutorial where they were introduced to the mechanics of the game: animals have life and have to eat to survive and the child can move animals to make them interact with other animals or plants and replenish their energy (cf. Figures S3.f,g). The teacher was sitting with the child through these steps to provide clarification if needed. After this short tutorial, the teacher sat away from the child to supervise the robot if required while the child completed two rounds of the game where the robot could provide feedback and advices depending on the condition they were in (cf. Figure S3.h-k). Following these initial rounds of the game, the child completed a mid-test before playing another two rounds of the game and completing a last post-test to conclude the study. Figure S3 presents examples of screenshot of the Sandtray throughout the interaction. In all condition, the robot verbally guided the child through the study by explaining them what they were expected to do in the different phases. The differences of behaviour between conditions happened only during the game, where the robot could be passive, supervised or autonomous.

Implementation The robot is controlled using the architecture presented in Figure 7 with all the nodes communicating together using the Robot Operating System (ROS) (*22*). The teacher interface runs on a separate tablet and is used only for the supervised condition. All the other nodes run on the large touchscreen computer displaying the game interface which is used to guide the child through the study and presents the game rounds and the tests. The default robot

behaviour is simply reading the instruction on the screen, following the child's face and swaying lightly.

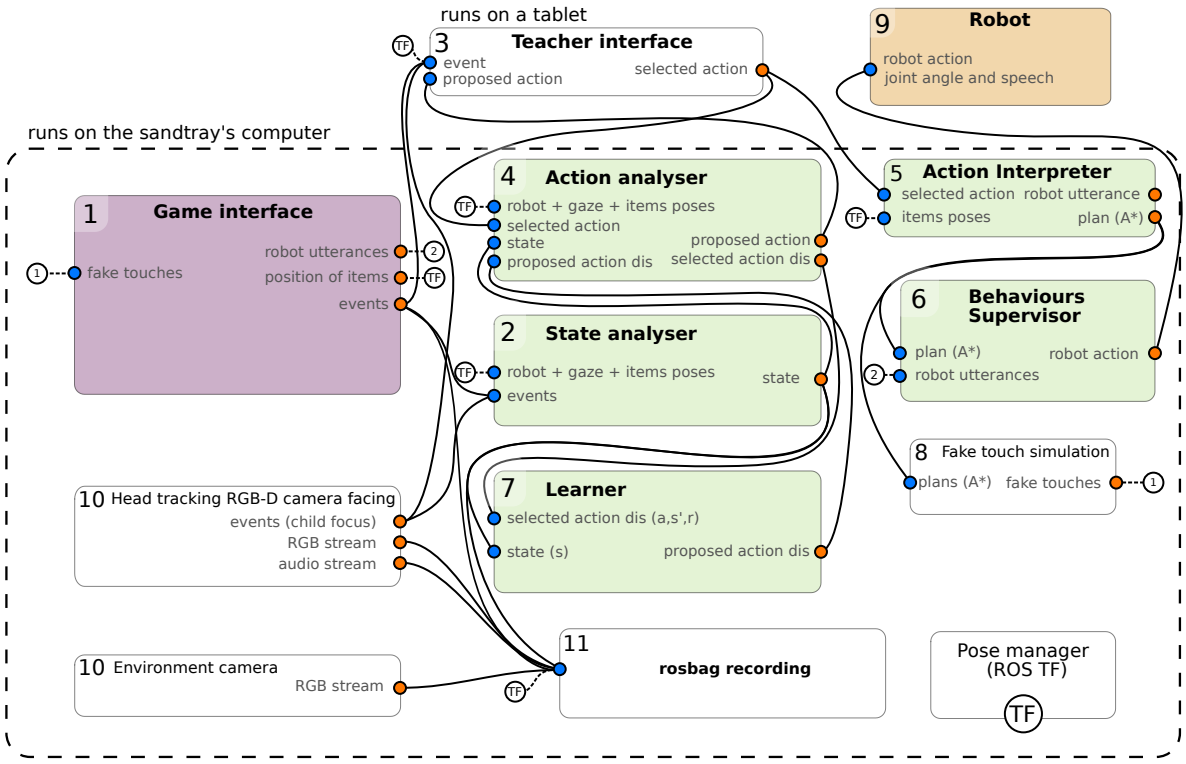

Figure 7: Simplified schematics of the architecture used to control the robot, the different nodes communicate using ROS. A game (1) runs on a touchscreen between the child and the robot. (2) analyses the state of the game using inputs from the game and the camera. (3) is an interface running on a tablet and used by the teacher to control and teach the robot. (4) communicates actions between the interface (3) and the learner (7). (5) translates teacher's actions into robotic commands used by (6) and (8) and executed by the robot (9). Finally, (7) is the learning algorithm which defines a policy based on the state perceived and the previous actions selected by the teacher, their substates and their feedback on propositions.

To support the children during the game rounds, the robot has access to 655 actions consisting on moving animals in relation to others on the screen (by pointing to an object and moving it on the screen), asking the child to focus on some items of the game (by pointing to them and uttering a predefined sentence) and providing social prompts and feedback such as reminding the rules and providing encouragements or congratulations. The robot's policy in the game

consists in a mapping between these actions and a representation of the state defined in a 210 dimensions vector with values ranging from 0 to 1 and corresponding features describing the state of the game (animal's energy, distance between items) and of the interaction (how long it has been since the child or the robot touched items, when was the last action executed by the robot...).

In the supervised condition, the teacher uses an interface running on a tablet and replicating the graphics of the game (with the position of the animals), but with additional buttons to select actions for the robot to execute. Our algorithm, adapted from (*14*), maps actions selected by the teacher to a *substate* (s'), a sliced version of the 210-dimension state using a variation of the Nearest Neighbours approach. In this new algorithms, instances in memory are only defined on a substate instead of the full state (n' dimensions of the state have a value, while the others, not relevant to the current action, are left as 'wild cards'). These substates are defined on S' state subspaces, corresponding to sliced versions of the state space (with $S' \subset S$). This slicing is carried out by keeping only the dimensions relevant to a set of features defined by the teacher (i.e. selected on the tablet). This allows the algorithm to consider only the dimensions of the state relevant to each action when computing the distance between instances and the current state. Consequently, this algorithm can profit from having access to a large number of state dimensions without suffering from the 'Curse of dimensionality' (*23*), thus potentially learning quickly complex behaviours. Additionally, each instance in memory possesses a reward value (r) which allows the algorithm to avoid undesired actions (the ones with a negative reward). In summary, instances are defined as tuples: action - substate - reward (a, s', r).

This learning algorithm can propose actions to the teacher that are executed after a short delay if the teacher does not cancel them. Using the interface the teacher can accept (rewarding positively and executing) proposed actions or refuse them (pre-empting the execution of an action and assigning it a negative reward). Additionally, they can select actions for the robot to

execute. Figure 8 shows the flowchart of the action selection process allowing mixed initiative between the teacher and the robot.

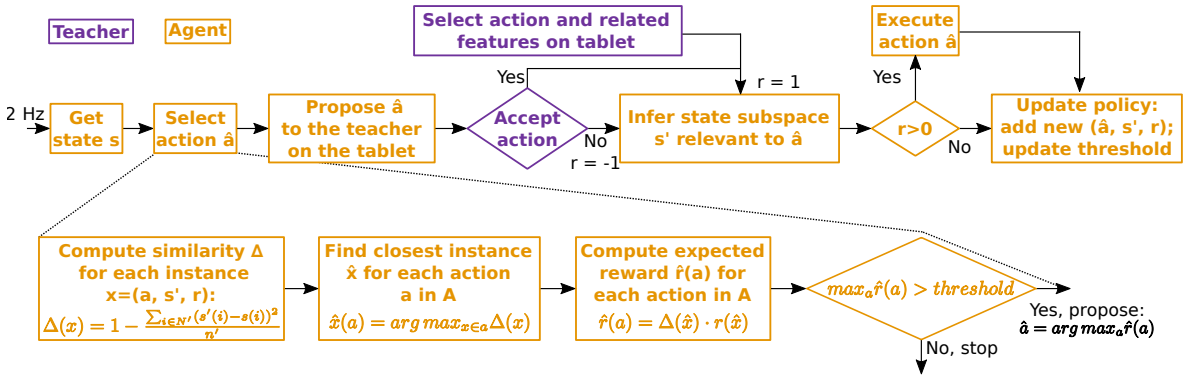

Figure 8: Flowchart of the action selection. Mixed-initiative control is achieved via a combination of actions selected by the teacher, propositions from the robot and corrections of propositions by the teacher. The algorithm uses instances x, corresponding to a tuple: action a , substate s' and reward r. s' is defined on S' with $S' \subset S$ and N the set of the indexes of the n' selected dimensions of s'.

The algorithm itself does not take time into account. However, as dimensions of the state are time dependant (using exponential decreases since events), temporal effects can be captured by the learning algorithm.

In the autonomous condition, the interface used by the teacher is simply replaced by a node automatically accepting propositions after a short delay, thus applying the policy learnt in the supervised condition.

Metrics To address the hypotheses, we collected multiple metrics on both interactions (teacher-robot and robot-child). First, we recorded the actions executed by the robot in the supervised and autonomous conditions to characterise the two policies. Second, we collected two groups of metrics to evaluate the application interaction: the learning metrics (corresponding to the child's performance during the tests) and the game metrics (corresponding to the child's behaviour within the game rounds). And finally, in the supervised condition, we recorded the

origin of the actions executed by the robot (teacher vs algorithm) and the outcome of the proposed actions (executed vs refused).

During the game, the robot had access to 655 actions, which can be divided into seven categories: drawing attention, moving close, moving away, moving to, congratulation, encouragement and reminding rules. Due to this high number of actions, the breadth of the state space (210 dimension) and the complex interdependence between actions and states, precisely characterising a whole policy is non-tractable. Consequently, we used the number of actions executed for each category per child to characterise the policy executed by the robot in the active conditions (supervised and autonomous). While not perfectly representing the policy of each condition (e.g. the timing of actions is missing), this metric offers a proxy to compare these policies.

The children's knowledge about the food web was evaluated through a graph where children had to connect animals to their food. There was 25 correct connections and 95 incorrect ones. As the child could create as many connections as desired, the performance was defined as the number of correct connections above chance (for the total number of connection made during the test) divided by the maximum achievable performance. This resulted in a score bounded between -1 and 1.

For example, if a child made 5 good connections and 3 bad, their performance would be:

$$P = \frac{\#good - (\#good + \#bad) \cdot \frac{totalgood}{total}}{totalgood - totalgood \cdot \frac{totalgood}{total}} = \frac{5 - (5+3) \cdot \frac{25}{25+95}}{25 - 25 \cdot \frac{25}{25+95}} = 0.168 \quad (1)$$

The three tests (pre, mid and post interaction) resulted in three performance measures.

To account for initial differences in knowledge and the progressive difficulty to gain additional knowledge, we computed the learning gain as the difference between the final and initial knowledge divided by the 'progression margin': the difference between the maximum achievable performance and the initial performance ($Learning\ gain = \frac{P_{final} - P_{initial}}{1 - P_{initial}}$). This learning

gain indicates how much of the missing knowledge the child managed to gain from the game. Add ref, but I did not find a good one, could you direct me to a paper Tony or Paul please?[ES]

Additionally, different metrics were also gathered during the rounds of the game to characterise the children's behaviours:

- **Number of different eating interactions**: number of unique eating interactions between two items ([0,25]).
- **Interaction time**: Duration of game rounds, how long a round lasted until three animals ran out of energy (typical range 0.5 to 3 minutes).

As mentioned in the previous section, the children had to explore a food net with 25 good connections and 95 incorrect connections. Due to the imbalance between these numbers, more knowledge is acquired by discovering one of these 25 good connections rather than disproving one of the 95 incorrect ones. As such, we defined our first game metric as the number of different eating interactions children encountered during each game. An eating interaction happens when the child moves an animal to its food (or to a predator); and the number of different eating interactions represents how many different unique correct connections the child has discovered to during the game (multiple eating actions between the same animals would count only once). A game with a high number of different eating interactions represents a game where the child had the opportunity to learn many correct connections between animals. Consequently, by increasing this number, the children would be exposed to more learning items which should help them perform better on the tests. For simplicity, we termed this metric 'exposure to learning items' as it encompasses how much knowledge a child has been exposed to in one round of the game. We would expect that an active robot would be able to guide the child towards these correct connections, allowing the child to reach a higher exposure, which would lead to more gain from the game and better overall learning.

On the other hand, the interaction time reached in the game provide information about the children's performance in the task (keeping the animals alive as long as possible) and their engagement. A child disengaged with the game would not play seriously and would finish the interaction earlier. We expect that an active robot would encourage and support the child in the learning game and allow them to reach better scores in these engagement and performance metrics.

Statistical Analysis To demonstrate the presence or the absence of effects, we used bayesian statistics on the data. As such the Bayes factor B_{10} is reported and represents how much of the variance on the metric is explained by a parameter (if $B_{10} < 1/3$ there is no impact, if $B_{10} > 3$ the impact is strong, and if $1/3 < B_{10} < 3$ the results are inconclusive (*24, 25*)). We analysed the results using the JASP software (*26*). We used a Bayesian mixed ANOVA as an omnibus test to explore the impact of the condition and the repetition on the metrics. Additional post-hoc tests used a Bayesian Repeated-Measure ANOVA comparing the conditions one by one and fixing the prior probability to 0.5 to correct for multiple testing. The left graphs are violin plots featuring the kernel density estimation of the distribution and right graphs present the mean and the 95% Confidence Intervals.

References

1. C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, M. Berlin, *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on* (IEEE, 2005), pp. 708–713.

2. R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction* (MIT press, 1998).

3. A. Billard, S. Calinon, R. Dillmann, S. Schaal, *Springer Handbook of Robotics* (Springer, 2008), pp. 1371–1394.

4. B. D. Argall, S. Chernova, M. Veloso, B. Browning, *Robotics and Autonomous Systems* **57**, 469 (2009).

5. P. Liu, D. F. Glas, T. Kanda, H. Ishiguro, N. Hagita, *23rd IEEE International Symposium On Robot and Human Interactive Communication (RO-MAN)* (2014), pp. 961–968.

6. P. Sequeira, *et al.*, *The Eleventh ACM/IEEE International Conference on Human Robot Interation* (IEEE Press, 2016), pp. 197–204.

7. M. Clark-Turner, M. Begum, *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (ACM, 2018), pp. 372–372.

8. J. A. Fails, D. R. Olsen Jr, *Proceedings of the 8th International Conference on Intelligent User Interfaces* (ACM, 2003), pp. 39–45.

9. S. Amershi, M. Cakmak, W. B. Knox, T. Kulesza, *AI Magazine* **35**, 105 (2014).

10. W. B. Knox, P. Stone, *Proceedings of the Fifth International Conference on Knowledge Capture* (ACM, 2009), pp. 9–16.

11. A. L. Thomaz, C. Breazeal, *Artificial Intelligence* **172**, 716 (2008).

12. T. L. Sanders, T. Wixon, K. E. Schafer, J. Y. Chen, P. Hancock, *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2014 IEEE International Inter-Disciplinary Conference on* (IEEE, 2014), pp. 156–159.

13. S. Chernova, M. Veloso, *Journal of Artificial Intelligence Research* **34** (2009).

14. E. Senft, S. Lemaignan, P. Baxter, T. Belpaeme, *Proceedings of the Artificial Intelligence for Human-Robot Interaction Symposium, at AAAI Fall Symposium Series* (2017).

15. D. Leyzberg, S. Spaulding, B. Scassellati, *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (ACM, 2014), pp. 423–430.

16. I. Leite, G. Castellano, A. Pereira, C. Martinho, A. Paiva, *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (ACM, 2012), pp. 367–374.

17. E. Senft, P. Baxter, J. Kennedy, S. Lemaignan, T. Belpaeme, *Pattern Recognition Letters* **99**, 77 (2017).

18. E. Senft, P. Baxter, J. Kennedy, T. Belpaeme, *International Conference on Social Robotics* (Springer, 2015), pp. 603–612.

19. T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, F. Tanaka, *Science Robotics* **3**, eaat5954 (2018).

20. T. Schodde, K. Bergmann, S. Kopp, *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (ACM, 2017), pp. 128–136.

21. P. Baxter, R. Wood, T. Belpaeme, *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (2012), pp. 105–106.

22. M. Quigley, *et al.*, *ICRA Workshop on Open Source Software* (Kobe, Japan, 2009), vol. 3, p. 5.

23. R. Bellman, *Dynamic Programming* (Princeton University Press, Princeton, 1957).

24. H. Jeffreys, *The Theory of Probability* (OUP Oxford, 1998).

25. Z. Dienes, *Perspectives on Psychological Science* **6**, 274 (2011).

26. JASP Team, JASP (Version 0.8.6)[Computer Software] (2018).

Acknowledgments

This work was supported by the EU FP7 DREAM project (grant no. 611391), the EU H2020 Marie Sklodowska-Curie Actions project DoRoThy (grant 657227) and the EU H2020 L2TOR project (grant 688014).

Raw data consist on rosbags and csv files, as such, preprocessed data, the script required to generate the graphs and JASP file for the statistical analysis can be found at `https://github.com/emmanuel-senft/data-foodchain`.

Table 1: Example of events during the first minute of the first round of the interaction with the 23rd child in the supervised condition. Lines in blue represent propositions from and the robot and orange the reaction from the teacher. ('mvc' is the abbreviation of the move close action)

Time	Event	Time	Event
4.1	childtouch *frog*	34.4	childtouch *wolf*
4.3	failinteraction *frog wheat-3*	34.7	robot proposes remind rules
4.9	animaleats *frog fly*	35.0	animaleats *wolf mouse*
5.8	childrelease *frog*	36.0	teacher selects wait
6.6	robot proposes congrats	36.0	animaleats *wolf mouse*
7.6	childtouch *fly*	37.2	childrelease *wolf*
7.6	teacher selects wait	37.7	childtouch *grasshopper*
8.0	animaleats *fly apple-4*	38.3	robot proposes congrats
8.3	childrelease *fly*	41.8	reset
9.1	teacher selects congrats	42.1	failinteraction *grasshopper apple-1*
9.1	childtouch *frog*	42.7	childrelease *grasshopper*
10.3	childrelease *frog*	42.7	failinteraction *grasshopper apple-1*
10.8	childtouch *frog*	44.4	teacher selects mvc *mouse wheat-1*
11.2	animaleats *frog fly*	44.6	robottouch *mouse*
12.4	failinteraction *frog apple-2*	44.7	childtouch *butterfly*
12.5	animaleats *frog fly*	45.1	failinteraction *butterfly wheat-2*
13.2	childrelease *frog*	45.6	childrelease *wheat-1*
14.2	childtouch *fly*	45.6	robotrelease *mouse*
14.5	animaleats *fly apple-2*	45.7	robottouch *mouse*
14.6	robot proposes encouragement	48.9	robotrelease *mouse*
15.0	childrelease *fly*	49.3	childtouch *butterfly*
15.4	animaleats *fly apple-3*	49.3	failinteraction *butterfly wheat-1*
16.9	teacher selects encouragement	49.6	childrelease *butterfly*
18.2	childtouch *snake*	50.0	childtouch *mouse*
18.4	failinteraction *snake wheat-3*	50.3	animaleats *mouse wheat-1*
18.7	animaleats *snake bird*	51.0	childrelease *mouse*
19.6	animaleats *snake bird*	51.1	animaleats *mouse wheat-2*
20.5	childrelease *snake*	51.4	robot proposes congrats
20.6	failinteraction *snake wheat-4*	52.3	teacher selects congrats
20.6	robot proposes congrats	52.9	childtouch *snake*
20.9	childtouch *eagle*	52.9	failinteraction *snake wheat-3*
21.1	animaleats *eagle bird*	53.2	childrelease *snake*
22.0	animaleats *eagle bird*	53.5	childtouch *mouse*
22.4	childrelease *eagle*	53.6	animaleats *mouse wheat-3*
23.3	animaldead *bird*	54.4	robot proposes congrats
23.4	teacher selects mvc *dragonfly fly*	54.5	animaleats *mouse wheat-4*
23.6	robottouch *dragonfly*	55.0	childrelease *mouse*
26.9	robotrelease *dragonfly*	55.6	childtouch *dragonfly*
27.7	childtouch *fly*	56.1	teacher selects wait
28.0	childrelease *fly*	56.8	failinteraction *dragonfly apple-1*
28.4	childtouch *dragonfly*	57.3	childrelease *dragonfly*
28.6	failinteraction *dragonfly apple-1*	57.5	failinteraction *dragonfly apple-1*
29.1	childrelease *dragonfly*	58.6	childtouch *grasshopper*
29.4	failinteraction *dragonfly apple-1*	58.6	failinteraction *grasshopper apple-1*
30.3	childtouch *dragonfly*	58.8	childrelease undefined
30.3	failinteraction *dragonfly apple-1*	59.1	childtouch *dragonfly*
30.7	robot proposes encouragement	59.1	failinteraction *dragonfly apple-1*
31.0	failinteraction *dragonfly apple-1*	59.2	failinteraction *grasshopper apple-1*
31.8	teacher selects wait	59.9	failinteraction *dragonfly apple-1*
32.5	childrelease *dragonfly*	60.3	childrelease *dragonfly*

Recognizing Human Internal States: A Conceptor-Based Approach

***Abstract*—The past few decades has seen increased interest in the application of social robots to interventions for Autism Spectrum Disorder as behavioural coaches [4]. We consider that robots embedded in therapies and interventions could also provide quantitative diagnostic information by observing patient behaviours. The social nature of ASD symptoms means that, to achieve this, robots need to be able to recognize the internal states their human interaction partners are experiencing, e.g. states of confusion, engagement etc. Approaching this problem can be broken down into two questions: (1) what information, accessible to robots, can be used to recognize internal states, and (2) how can a system classify internal states such that it allows for sufficiently detailed diagnostic information? In this paper we discuss these two questions in depth and propose a novel, conceptor-based classifier. We report the initial results of this system in a proof-of-concept study and outline plans for future work.**

***Index Terms*—Internal States, Engagement, Conceptors, Socially Interactive Robots, Recognition**

I. Introduction

The development of socially interactive robots has inspired research into various applications for these tools. One application is in therapy and care, where robots can be used to provide daily support to patients, and as tools to augment interventions and provide quantitative data for clinicians [1]. We specifically consider the use of robots in interventions for children with Autism Spectrum Disorder (ASD). The Diagnostic and Statistical Manual of Mental Disorders (DSM-V) defines ASD as a neuro-developmental disorder characterized by persistent deficits in social communication and interaction, and restricted or repetitive behaviours and interests [2]. Diagnosing ASD involves the subjective interpretations by experts of observations of a child's behaviour made by clinicians and caregivers [3]. This subjectivity, and the clinical heterogeneity which is typical between ASD cases [4], means that the diagnostic process could be improved through the use of more quantitative, objective measures of child behaviour. This can be achieved using behaviour classification systems.

Developing an artificial system to recognize ASD symptoms is not a straight-forward task due to the social nature of ASD. This is because correct classification of social and interaction behaviour often requires the ability to infer the internal-states (e.g. intentions, emotions) of the observed individual. For example, identifying when a child fails to ask for comfort when needed (a symptom of ASD [2]) requires that the observer recognize that the child is experiencing a negative internal state. However, endowing robots with this skill would provide numerous benefits for ASD interventions. For instance, if an intervention involves regular interaction with a social robot, it would be useful to have the robot able to report quantitative diagnostic information. Firstly, clinicians could use this information to track their patient's progress through the intervention, or to support their initial diagnostic decision. Secondly, the robot itself could use internal-state and diagnostic information to autonomously decide on appropriate behaviours to perform.

In approaching the problem of developing artificial systems able to recognize human internal states, there are two key questions which must be addressed: (1) what internal state information is available in behaviours which can be assessed and quantified by artificial systems, and (2) how can these states be represented by a classification system to provide both detailed assessments and flexible behavioural responses from a social robot. The rest of this paper discusses possible answers to these questions in the context of quantifying the diagnostic behaviours of children with ASD. We present two studies carried out as a proof-of-concept to demonstrate that the internal state of task engagement could be classified based on observable human movement information, and that this classification could be done by a system able to represent internal states as points along a continuous dimension. The logic behind our choice of internal state and its desired representation is described, where relevant, in the introductions to each experiment.

II. Experiment 1

Whilst most ASD symptoms cannot be described as wholly overt, many have been linked with directly observable behaviours. For example, motor skills have been shown to be predictive of social communication skills for children with ASD [5]. Additionally, an increased tendency to orient towards non-social contingencies rather than biological motion is indicative of ASD [6]. These and other studies have linked movement and gaze behaviours to a number of ASD characteristics. Movement and gaze information can be measured or estimated by observing body movements or poses, which can be easily made accessible to artificial systems, e.g. by converting the position of an individual's joints to coordinates in space for each time-point. Consequently, we argue that this type of information can be useful for social robots designed to make inferences about diagnostic status and human internal states. Before a classifier could be implemented, however, we first needed to verify that the internal state of interest (task engagement) was recognizable from the movement information available in our data set.

For this proof-of-concept study, the desired data set was defined as one which contained the movement information of humans experiencing, but not intentionally communicating, different levels of a non-emotional internal state. To ensure that the internal state was not being communicated we decided that the subject should not be interacting with another human. The decision that the internal state should not be an emotion was made because we argue that the recognition of non-emotional internal states (e.g. task engagement, experienced difficulty) is more valuable for robots designed to interact in an intervention setting. For example, if a robot is able to recognize a child's engagement with an intervention task, the robot can decide when it is appropriate to provide the child with prompts or encouragement. However, the ensured presence of a human (i.e. the clinician), and the vulnerability of the target population (i.e. children) means that it is arguably safer and more ethical to provide human interaction in the event of emotional distress or discomfort.

With these considerations in mind, the data set selected for this experiment was taken from the openly available PInSoRo data set [7][1]. This data set comprises videos of child-robot pairs interacting with each other and a touch-screen table-top (the sand-tray). We argue that these videos meet the requirements of showing humans experiencing internal states which could be described along a continuum (i.e. engagement with the touch-screen) which were not being actively communicated (i.e. due to the lack of a human interaction partner). The videos have been annotated for a number of behaviors including whether the child was engaged in "goal oriented", "aimless" or "no" play. We believe these annotations are analogous to "high", "intermediate" and "low" levels of task engagement respectively. A preliminary study was designed to validate this assumption.

A. Method

1) Participants: Five participants (students and employees) were recruited from the University of Plymouth's School of Computing, Electronics and Mathematics on a volunteer basis. Demographic information was not collected.

2) Materials: A total of forty-five video clips were extracted from the data set for this study. We selected fifteen clips with the annotation "goal-oriented play", fifteen with the annotation "aimless play" and fifteen with the annotation "no play". Clip lengths ranged from 12-30 seconds.

After clips were selected we extracted both the full visual scene versions and the movement-alone versions. The movement-alone versions were processed such that they depicted the children's joint-points, connected by coloured lines, against a black background. These videos act as visual representations of the data used as input for the conceptor-based system in that they depict only movement and pose information by showing the position of the child's body in each frame.

[1]https://freeplay-sandbox.github.io

3) Procedure: For each participant the experiment was conducted over two days. Participants watched the full visual scene videos on the first day and were then asked to return the next day when they would watch the movement-alone videos. Participants all received the following instructions before beginning the experiment:

You're about to watch several videos of children interacting with a touch-screen table-top. The children were able to either play a specific game on the touch-screen, or to do whatever they want. After each clip you will be asked to judge the child's level of task engagement.

Participants were then given the opportunity to ask any questions they may have had and were instructed about their right to withdraw before beginning the experiment.

This study was created using JSPsych and presented on a desktop computer. Participants were positioned a comfortable distance away from the screen where they could still reach the keyboard and mouse to provide responses. At the beginning of the experiment, the instructions were reiterated. Participants were then presented with a consent form within the experiment script and given two response options. If participants selected the "I consent" option, the experiment proceeded as normal. If participants selected "I do not consent" the experiment was terminated. Participants then viewed nine of each type of clip (a total of twenty-seven clips) presented in a random order. Following each clip, participants were presented with the question "*How engaged was the child with their task on the touch screen table-top?*". This question was accompanied by a 7-point Likert scale ranging from 1 = "Not at all Engaged" to 7 = "Highly Engaged". Participants used this scale to report how engaged they thought the child in the clip had been and then continued on to the next clip.

At the end of the experiment on the first day, participants were given the opportunity to ask any questions they had and were asked to return the next day to complete the second half. On the second day, the experiment proceeded in the same way except participants were shown the movement-alone videos instead of the full visual scene videos. Each participant saw the same twenty-seven clips in both sessions. At the end of the second session participants were fully debriefed on the nature and purpose of the study and were thanked for their participation. Each session took approximately 10-15 minutes to complete.

B. Results

The following analyses were run using RStudio.

1) Inter-Rater Agreement: The data were analyzed in two main ways. We firstly examined inter-rater agreement by calculating Krippendorff's alpha for the responses. We initially checked whether participants gave similar responses for each of the three types of videos. To do this, Krippendorff's alpha was calculated for responses to all of the videos of each type. The alpha scores have been interpreted in terms of the benchmarks outlined by Landis and Koch [8]. Responses showed "fair" agreement for the goal-oriented (high engagement)

clips (Krippendorff's alpha = 0.269) and the no-play (low engagement) clips (Krippendorff's alpha = 0.267). Responses for aimless (intermediate engagement) clips showed "slight" agreement (Krippendorff's alpha = 0.171). The low levels of agreement can partially be explained by the fact that there were very few raters (2-4) per clip. As such we did not expect perfect levels of agreement and argue that the levels obtained suggest a sufficient degree of similarity in participants' ratings.

We then examined whether participants had higher agreement when viewing the full visual scene clips compared to the movement-alone clips for each clip type. The results of this analysis are reported in Table 1. For the goal-oriented and no-play clips, participants tended to show similar levels of agreement in each condition. However, for the aimless clips, participants demonstrated poor agreement when viewing the movement-alone clips.

TABLE I
TABLE OF INTER-RATER AGREEMENT SCORES FOR RESPONSES TO EACH CLIP-TYPE IN EACH CONDITION

Clip Type	Krippendorff's Alpha (3 d.p.)	
	Full Scene	movement-alone
Goal Oriented	0.382 (fair)	0.368 (fair)
Aimless	0.247 (fair)	-0.022 (poor)
No Play	0.126 (slight)	0.202 (fair)

2) Ratings: The second set of analyses looked at the how participants rated each type of video. Overall mean rating was 4.81 (SD = 1.25) for goal-oriented clips, 4.16 (SD = 1.52) for aimless clips, and 2.43 (SD = 1.54) for no-play clips. An ANOVA revealed a significant main effect of clip-type on ratings ($F(2,267)=64.99$, $p<0.001$). Importantly, a *post hoc* Tukey test revealed significant differences between all conditions (Tukeys HSD: all differences >0.6, all ps <0.007; see Table 2).

TABLE II
TABLE OF RESULTS FOR POST HOC TUKEY'S HONEST SIGNIFICANT DIFFERENCE TEST.

Comparison	Difference	Significance (p adj)
Goal Oriented − Aimless	0.656	$p = 0.007$
Goal Oriented − No Play	2.348	$p < 0.001$
Aimless − No Play	1.722	$p < 0.001$

These results demonstrate that participants rated the clips in terms of engagement such that goal-oriented clips showed the highest levels of engagement, no-play clips showed the lowest levels, and aimless clips fell in-between these two extremes. Consequently, we feel our assumption that these annotations reflect different levels of engagement is sufficiently supported for these data to be used to train and test a conceptor-based classifier to recognize engagement based on observable behaviour. The remainder of this paper describes the design and initial tests of such a classifier.

III. EXPERIMENT 2

In addressing the second question of how to represent internal states, we consider that ASD diagnosis involves ranking behaviours in terms of severity [9]. In this way, behaviours important to ASD diagnosis can be thought of as lying along a continuum of severity. To emulate this we therefore want a classification technique which can identify different 'levels' along a continuous dimension. This can be achieved using classical machine learning techniques by training a classifier on examples of each severity level. However, obtaining a large enough training data set for this would be very time-consuming and difficult, owing to the need to have expert commentators provide a severity label for each example. We therefore require a method which can learn several classification categories for each behaviour of interest, using a limited training data set. One approach which is suited to this task is conceptors [10].

Conceptors are neuro-computational mechanisms that can be used for learning a large number of dynamical patterns based on learned prototypical extremes [10]. This approach assumes that there is a continuum underlying the behavior. New patterns can be generated by combining and morphing the learned extremes. As such, we argue that conceptors may be appropriate for classifying human internal states. The second study described here tested this hypothesis by designing a conceptor-based system to recognize task engagement from observable human movements.

A. Method

1) Materials: The data set for this study was again taken from the PInSoRo data set. All of the clips annotated with the labels "goal-oriented play" (high engagement) and "no play" (low engagement) were extracted (total of 354 clips). Clips were preprocessed such that the xyz coordinates of the child's joints in each frame were taken as the input for the conceptor-based classifier. A subset of "high" (62 clips) and "low" (115 clips) engagement clips made up the training data set. The remaining 177 clips made up the test data set.

2) Conceptor-Based Classifier: The conceptor-based approach is based on a key dynamical phenomenon in Recurrent Neural Networks; "if a 'reservoir' is driven by a pattern, the entrained network states are confined to a linear subspace of network state space which is characteristic of the pattern" [10]. In this way the dynamics of a pattern (in our case an overt behavior for a classifiable activity like engagement) will occupy different regions of the state space, and they can be encoded in a conceptor. A conceptor (C_j) acts as a map associated with a pattern (p_j). To build a conceptor-based classifier we computed J conceptors, one for each class in our classifier. To obtain the conceptors an echo state network (ESN) was first created with an input layer of K input units and a hidden layer reservoir of N neurons. For each class the network will be driven, independently, with all training samples s_j^m in each class j, according to the ESN state update equation:

$$x(n+1) = \tanh(W \cdot x(n) + W^{in} \cdot p(n+1) + b) \qquad (1)$$

This yielded a set of network states $X_j = [x(1) \dots x(t)]$ where t is the number of time-steps in s_j from which a state

TABLE III
PREDICTING INTERNAL STATES WITH CONCEPTORS.

Algorithm: Conceptor-based classification.
Input: A test sample s belonging to one a class j.

1) Take a sample s from the test set.
2) Drive the reservoir with sample s to obtain a state vector $z = [x(1) \cdots x(n)]$, where n is the # of steps in s.
3) For each Conceptor C_j compute $h(j) = z^T C_j z$, a "positive evidence" quantity of z belonging to class j.
4) Collect each evidence $h(j)$ into a $j-$dimensional classification hypothesis vector $h^+ = \{h(1) \cdots h(j)\}$.
5) Classify s as belonging to class j from $j = argmax(h^+)$.
6) **END**

Output: Class sample s belongs to.

correlation matrix $R_j = X_j X_j^T / M_j$ is obtained, where M_j is the total number of samples for class j. Next we computed conceptor C_j through the equation:

$$C(R, \alpha) = R(R + \alpha^{-2})^{-1} \quad (2)$$

Where R is a correlation matrix and $\alpha \in (0, \infty)$ in an "apperture" parameter. For more see [10].

Once we computed a conceptor matrix for each class we were able to classify a new sample s from the test set by feeding it into the ESN reservoir to obtain a new state vector $z = [x(1) \dots x(n)]$. then, for each conceptor, the "positive evidence" quantity $z^T C_j z$ was computed. This led to a classification by deciding for $j = argmax(z^T C_j z)$ as the class j to which the sample s belongs. The procedure for the conceptor-based classifier is summarize in Table III-A2.

B. Results

The resultant conceptors were tested using previously unseen samples from the high and low engagement categories. The results of this test are shown in Figure 1. Performance is above chance for both classes (high engagement: 60%, low engagement: 75%).

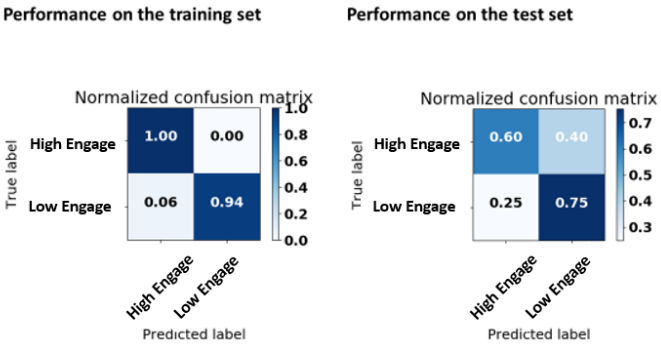

Fig. 1. Confusion matrices showing classification performance of trained conceptors on training data (left) and test data (right).

IV. CONCLUSIONS

This study demonstrates that it is possible to train a conceptor-based system, on real non-periodic data, to classify between high and low engagement based on observable human behavior. The conceptor-based system successfully learned to recognize high and low engagement from observable human movement. Future work will construct new conceptors by linearly combining these learned conceptors. We will then test whether these new conceptors can be used to recognize intermediate levels of engagement identified in the PInSoRo data set.

If new conceptors can be generated, this method will show promise for use in providing diagnostic information for clinicians assessing children with ASD. The ability to interpolate between extremes along a continuum means that such a system could be trained on a smaller dataset, whilst still achieving a high level of detail through the generation of multiple intermediate classification categories.

ACKNOWLEDGMENT

This work is part of the EU FP7 project DREAM project (www.dream2020.eu), funded by the European Commission (grant no. 611391)

REFERENCES

[1] H. M. Van der Loos, D. J. Reinkensmeyer, AND E. Guglielmelli, "Rehabilitation and health care robotics." In Springer handbook of robotics, pp. 1685-1728, 2016.
[2] American Psychiatric Association, "Diagnostic and statistical manual of mental disorders: DSM-5." Autor, Washington DC, 5th ed, 2013.
[3] C. L. Rogers, L. Goddard, E. L. Hill, L. A. Henry, and L. Crane, "Experiences of diagnosing autism spectrum disorder: a survey of professionals in the United Kingdom." Autism, vol. 20(7), pp. 820–831, 2016.
[4] B. Scassellati, H. Admoni, and M. Matari, "Robots for use in autism research." Annual review of biomedical engineering, vol. 14, pp. 275–294, 2012.
[5] J. Bradshaw, C. Klaiman, S. Gillespie, N. Brane, M. Lewis, and C. Saulnier, "Walking Ability is Associated with Social Communication Skills in Infants at High Risk for Autism Spectrum Disorder." Infancy, 2018.
[6] A. Klin, D. J. Lin, P. Gorrindo, G. Ramsay and W. Jones, "Two-year-olds with autism orient to non-social contingencies rather than biological motion." Nature, vol. 459(72440, pp. 257–261, 2009.
[7] S. Lemaignan, C. Edmunds, E. Senft, and T. Belpaeme, "The Free-play Sandbox: a Methodology for the Evaluation of Social Robotics and a data set of Social Interactions." arXiv preprint arXiv:1712.02421, 2017.
[8] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data." Biometrics, pp. 159–174, 1977.
[9] C. Lord, S. Risi, L. Lambrecht, E. H. Cook Jr, B. L. Leventhal, P. C. DiLavore, A. Pickles and M. Rutter, "The Autism Diagnostic Observation ScheduleGeneric: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism." Journal of Autism and Developmental Disorders, vol. 30(3), 2000.
[10] H. Jaeger, "Controlling recurrent neural networks by conceptors." arXiv preprint arXiv:1403.3369, 2014.