
Expert-Informed Design and Automation of Persuasive, Socially Assistive Robots

By

KATIE JAYNE WINKLE



BRISTOL ROBOTICS LABORATORY
UNIVERSITY OF THE WEST OF ENGLAND AND UNIVERSITY OF BRISTOL

A dissertation completed at the FARSCOPE Centre for Doctoral Training in Robotics and
Autonomous Systems.

JUNE 2020

Word count: 63,275

ABSTRACT

Socially assistive robots primarily provide useful functionality through their social interactions with user(s). An example application, used to ground work throughout this thesis, is using a social robot to guide users through exercise sessions. Initial works have demonstrated that interactions with a social robot can improve engagement with exercise, and that an embodied social robot is more effective for this than the equivalent virtual avatar. However, many questions remain regarding the design and automation of socially assistive robot behaviours for this purpose. This thesis identifies and practically works through a number of these questions in pursuit of one ultimate goal: the meaningful, real world deployment of a fully autonomous, socially assistive robot.

The work takes an expert-informed approach, looking to learn from human experts in socially assistive interactions and explore how their expert knowledge can be reflected in the design and automation of social robot behaviours. It is taking this approach that leads to the notion of socially assistive robots needing to be *persuasive* in order to be effective, but also identifies the difficulty in automating such complex, socially intelligent behaviour. The ethical implications of designing persuasive robot behaviours are also *practically* considered; with reference to a published standard on ethical robot design.

The work culminates with use of a state of the art, interactive machine learning approach to have an expert fitness instructor *train* a robot ‘fitness coach’, deployed in a university gym, as it guides participants through an NHS exercise programme. After a total of 151 training sessions across 10 participants, the robot successfully ran 32 sessions autonomously. The results demonstrated that autonomous behaviour was generally comparable to that of the robot when controlled/supervised by the fitness instructor, and that overall, the robot played an important role in keeping participants motivated through the exercise programme.

ACKNOWLEDGEMENTS

Firstly, a huge thank you to everyone involved in running the FARSCOPE Centre for Doctoral Training that made this possible. Being part of FARSCOPE has simply always made things just that little bit [easier / less daunting / more fun]. To the other FARSCOPERS and particularly everyone in my cohort who has been there to celebrate the highs and brush off the lows. To Chanelle for plying me with delicious home cooked food and for that all nighter we pulled in first year, completing our kinematics coursework the day before the deadline whilst listening to Smooth radio.

I am grateful also for the freedom that the FARSCOPE model affords to students with regards to proposing their own research projects. To this end, I must also say a heartfelt thank you to my supervisors Ailie, Ute, Praminda, and Paul. You have been there with me all of the way as I've jumped into everything from occupational therapy to persuasion psychology to the ethics of robot design. You've supported and encouraged this interdisciplinarity, guiding me where necessary and asking all the right questions whilst letting me really own the work and its direction.

To Séverin, essentially an unofficial supervisor, but also an excellent collaborator and ultimately a good friend. Thank you for being so excitable when it comes to interesting (and technically difficult!) research ideas, for sharing your positivity and energy when I was running low, and for helping me develop the technical skills needed to deliver on my ideas. Outside of work, thanks to you, Claire (and the kids!) for the dinners, chocolate and drinks that we've shared, especially when things have been difficult.

Thanks also to the other ECHOS researchers for being wonderful colleagues - whether delivering coffee during crazy study preparations (Alex) or telling me to go home at 5pm on a Friday (Manuel). More broadly, to everyone at the Bristol Robotics Lab who has ever stopped by my desk to say hi, check how I'm doing, or dish out various edible goodies.

Final thanks go to my family. To my Dad for always being interested in reading my latest paper. To Sam for reassuring me that it was *absolutely* valid to give up a very sensible graduate job in order to pursue this instead. To Jack for introducing me to all sorts of cool psychology literature and being an excellent conversation companion. And to my Mom, for being an absolute pillar of support throughout.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: DATE:

CONTENTS

	Page
List of Tables	xi
List of Figures	xiii
1 Introduction	1
1.1 Related Work	2
1.1.1 Mutual Shaping	5
1.2 Research Questions	6
1.3 Scope	6
1.3.1 Choice of Robot	7
1.3.2 Framing	7
1.4 Research Philosophy	9
1.5 Structure	10
2 A Study with Experts in Socially Assistive Interaction	13
2.1 Introduction	13
2.1.1 Related Work	14
2.1.2 Research Questions	16
2.2 Materials and Methods	16
2.2.1 Focus Groups	16
2.2.2 Interviews	23
2.2.3 Data Analysis	27
2.3 Findings	27
2.3.1 Use of SARs in Therapy (RQ1)	27
2.3.2 Measuring Engagement with Therapy (RQ2)	29
2.3.3 Therapists' Role in user Engagement (RQ3)	30
2.3.4 Personalised Approaches (RQ4)	32
2.4 Design Implications for Socially Assistive Robots	32
2.4.1 Improving Ease of Access	34
2.4.2 Improving Motivation	35

CONTENTS

2.4.3	Personalisation & Adaptability	36
2.4.4	Socially Assistive Robots Need Social Influence	38
2.5	Conclusion	38
3	Persuasion as a Model for Socially Assistive HRI	41
3.1	Introduction	41
3.1.1	Persuasion and Social Influence in HHI	43
3.1.2	Related Work	46
3.1.3	Ethical Considerations	48
3.1.4	Research Questions and Overview of Studies Presented	49
3.2	Study 1: ELM for Socially Assistive Robotics	51
3.2.1	Methodology	52
3.2.2	Experimental Measures	52
3.2.3	Experimental Conditions	54
3.2.4	Experimental Procedure	56
3.2.5	Results	58
3.2.6	Summary of Study 1 and Results	70
3.3	Study 2: Varying Suggested Expertise Source	71
3.3.1	Methodology	71
3.3.2	Experimental Conditions	73
3.3.3	Results	73
3.3.4	Summary of Study 2 and Results	77
3.4	Study 3: (More) Ethical Design of Social Dialogue	78
3.4.1	Methodology	78
3.4.2	Experimental Conditions	79
3.4.3	Results	80
3.4.4	Summary of Study 3 Results	84
3.5	Summary of Findings	86
3.5.1	Socially Persuasive Strategies on Robot Persuasiveness (RQ1)	86
3.5.2	Socially Persuasive Strategies on Perception of the Robot (RQ2)	86
3.5.3	Correlations Between Credibility/Likeability & Persuasiveness (RQ3)	87
3.5.4	Socially Persuasive Behaviour: Deception and Acceptability (RQ4)	87
3.5.5	Impact of (More) Ethical Dialogue on Persuasive Effectiveness (RQ5)	88
3.6	Discussion	88
3.6.1	Persuasion as a Model for Social, Assistive HRI	88
3.6.2	The Difficulty of Assessing Robot Credibility/Likeability	89
3.6.3	Designing Socially Persuasive Robots, <i>Ethically</i>	92
3.7	Conclusion	94

4 Creating an Autonomous, Socially Assistive Robot with Interactive Machine Learning	97
4.1 Introduction	97
4.1.1 Interactive Machine Learning for Automating SAR Behaviour	98
4.1.2 Couch to 5km: a Real World SAR Application	100
4.2 Technical Approach	100
4.2.1 Assumptions and Limitations	100
4.2.2 Modelling the Robot Action Space	103
4.2.3 Interaction Features and Input Space	105
4.2.4 Dual Learner System	106
4.2.5 IML System Architecture	106
4.2.6 Supervised Learning: Information Flow	113
4.3 Real World Deployment: Learning & Evaluation Study	115
4.3.1 Research Questions and Hypotheses	115
4.3.2 Gym Installation and Setup	116
4.3.3 Participants	116
4.3.4 Fitness Instructor	118
4.3.5 Conditions	118
4.3.6 Testing Schedule	119
4.3.7 Experimental Measures	120
4.3.8 Technical Limitations and Run-time Fixes	122
4.4 Findings	124
4.4.1 Example Session	124
4.4.2 Usability for Generating Appropriate Action Policies (RQ1)	126
4.4.3 Autonomous Robot Behaviour (RQ2)	137
4.4.4 Participant Experience of Couch to 5km with C25K Robot (RQ3)	148
4.5 Discussion	151
4.5.1 Successfully Demonstrating the Potential for SARs In-the-Wild	151
4.5.2 IML for Generating Autonomous SAR Behaviour	153
4.5.3 The Practicalities of IML as a Process for SAR Automation	154
4.6 Conclusion	155
5 Mutual Shaping in Design and Deployment of Socially Assistive Robots	157
5.1 Introduction	157
5.1.1 Methods for Mutual Shaping	158
5.1.2 Related Work	159
5.2 Extended Focus Groups for Mutual Shaping	162
5.2.1 How: Extended Focus Group Methodology	162
5.2.2 Why: Insightful Results Concerning real world Deployment	164

CONTENTS

5.2.3	Mutual Shaping: Impact on Participant Acceptance	169
5.3	IML as Participatory Design for Mutual Shaping	170
5.3.1	How: IML versus Heuristic Implementation of an Autonomous SAR	170
5.3.2	Why: Benefits of Process and Resulting Autonomous System	179
5.3.3	Mutual Shaping: Interactions Between the IML System, Fitness Instructor and Participants	191
5.4	Conclusion	194
5.4.1	Limitations	196
6	Conclusion	197
6.1	Getting Human, Domain Expert Knowledge into SARs (RQ1)	199
6.2	On the Value of Taking Expert-Informed and Mutual Shaping Approaches (RQ2) .	200
6.3	The Role of Humans in Socially Assistive HRI (RQ3)	202
6.4	Ethical Considerations	203
6.5	Limitations	204
6.6	Future Work	205
6.6.1	Variations on the Expert Role and Learning from the User	205
6.6.2	Alternative Machine Learning Approaches to Utilise Expert Input	206
6.7	Concluding Summary	207
A	Appendix A	209
B	Appendix B	221
C	Appendix C	229
D	Appendix D	255
Bibliography		263

LIST OF TABLES

TABLE	Page
2.1 Study with experts complete participant list	17
2.2 Focus group participant groupings	17
2.3 Focus group schedule	18
2.4 Interview schedule and topic guide	24
2.5 Patient traits that inform personalisation in therapy	33
2.6 Elements of therapy that are personalised to the patient	34
3.1 Bipolar adjectives for measuring credibility with a semantic difference scale	45
3.2 An overview of related HRI literature concerning SARs and/or persuasiveness	46
3.3 Overview of experimental studies on persuasive social robot behaviour	50
3.4 Robot dialogue employed for each Study 1 experimental condition	55
3.5 Summary of quantitative Study 1 results	71
3.6 Robot dialogue employed for each Study 2 experimental condition	74
3.7 Summary of quantitative Study 2 results	78
3.8 Robot dialogue employed for each Study 3 experimental condition	81
3.9 Summary of quantitative Study 3 results.	86
4.1 Participatory design activities undertaken for the C25K robot coach	101
4.2 List of C25K robot coach actions	104
4.3 Description of each C25K robot coach action-type	104
4.4 Example dialogue for each of the C25K robot coach speech-based actions	105
4.5 List of C25K robot coach inputs	106
4.6 Example C25K session testing schedule	120
4.7 Total number of experimental sessions conducted per condition, per participant	121
4.8 Total number of training data points collected	124
4.9 Example supervised session: actions suggested and/or executed	125
4.10 Subset of sessions used to compare supervised and autonomous robot behaviour	139
4.11 Fitness instructor notes for a subset of autonomous sessions	149
4.12 Qualitative participant data collected at end of study	152

LIST OF TABLES

5.1	Key elements of focus group methodology as applied to the study with therapists . . .	163
5.2	Pre/post-session participant robot acceptance comparison	170
5.3	Comparison of IML and Heuristic system action spaces/action space utilisation . . .	178
5.4	Comparison of input spaces for the IML and Heuristic systems.	178
5.5	Subset of sessions used to compare C25K robot behaviours	181
D.1	Fisher's exact test results comparing robot behaviours, within-participant	256
D.2	Fisher's exact test results comparing H1 robot behaviour across participants . . .	257
D.3	Fisher's exact test results comparing H2 robot behaviour across participants . . .	257
D.4	Fitness instructor notes from a subset of initial heuristic sessions	258
D.5	Fitness instructor notes for a the final heuristic sessions	259
D.6	Participant descriptions of robot, fitness instructor and researcher role	260
D.7	Participant descriptions of the IML and H robots	261

LIST OF FIGURES

FIGURE	Page
2.1 Study with experts focus group room layout	18
2.2 Focus group example ranking task output	21
2.3 Therapist descriptors and application of NHS categorisation tool for two service users	25
2.4 Two NHS Healthy Foundations segmentation tool personas	26
2.5 Valence of participant comments before/after focus groups	28
3.1 Age distribution for Study 1 participants	53
3.2 Study 1 experimental setup	57
3.3 Study 1 interaction pattern and wizard protocol	58
3.4 Study 1 objective user behaviour results (wrist turn repetitions across condition) . . .	59
3.5 Study 1 results concerning robot genuineness and deception across conditions . . .	61
3.6 Emergent themes from Study 1 on how participants described the robot	62
3.7 Emergent themes from Study 1 on what participants liked about the robot	63
3.8 Emergent themes from Study 1 on what participants disliked about the robot	64
3.9 Emergent themes from Study 1 on robot ‘genuineness’	65
3.10 Emergent themes from Study 1 on whether the robot was deceptive	67
3.11 Emergent themes from Study 1 on acceptability of the robot’s behaviour	68
3.12 Age distribution for Study 2 participants	72
3.13 Scene setup used for all videos produced for Studies 2 and 3	73
3.14 Study 2 results on perceived source of robot information	76
3.15 Study 2 results on which robot was most motivating/participants would rather work with	76
3.16 Emergent themes from Study 2 regarding participants’ preferred choice of robot . .	77
3.17 Age distribution for Study 3 participants	79
3.18 Study 3 results on which robot was most motivating/participants would rather work with	83
3.19 Emergent themes from Study 3 regarding participants’ preferred choice of robot . .	84
3.20 Study 3 results on robot deception	85
3.21 Emergent themes from Study 3 regarding robot deception	85

LIST OF FIGURES

4.1	Diagrammatic depiction of expert-in-the-loop, interactive machine learning	99
4.2	Overview of how the dual learning system works	107
4.3	The C25K coach and learning system architecture	107
4.4	Screenshots of the C25K teacher interface	108
4.5	The C25K teacher interface ‘in-action’ showing Learner suggestions	109
4.6	The C25K Robot-mounted tablet showing the ‘Check PRE’ action	110
4.7	Diagrammatic depiction of the experimental setup	116
4.8	Photographs of the system in use during an experimental session	117
4.9	Emoji-based scale used in post-session measures	122
4.10	Actions executed during supervised sessions	128
4.11	Use of action space in mixed vs pure run sessions	129
4.12	Number of actions used in mixed vs pure run sessions	130
4.13	Use of styles across participants for final common supervised session	131
4.14	Use of action-types across participants for final common supervised session	132
4.15	Total actions executed across all participants’ session 17	133
4.16	Cumulative sum of action types for two participants’ session 17	134
4.17	Cumulative sum of training actions collected as training data	135
4.18	Instructor response to Learner suggested task action styling	136
4.19	Overall instructor workload across all Phase 2 training sessions	138
4.20	Comparison of style distribution for supervised and autonomous sessions	140
4.21	Comparison of action distribution for supervised and autonomous sessions	141
4.22	Executed action styles across participants for autonomous sessions	143
4.23	Executed action-types across participants for autonomous sessions	144
4.24	Total actions executed across comparable participants’ autonomous sessions	145
4.25	Cumulative sum of action types for two participants’ autonomous session 23	146
4.26	Comparison of supervised vs autonomous post-session robot ratings	147
4.27	Coding of fitness instructor autonomous session notes	150
4.28	Participant responses to working with a robot again in the future	151
5.1	Example shift in participant acceptance	169
5.2	Participant acceptance questionnaire responses	171
5.3	Participant acceptance questionnaire mean scores	172
5.4	IML versus heuristic design processes for automating robot behaviour	173
5.5	Fitness instructor conducting mock C25K session with a researcher	174
5.6	Physical prototyping of the teaching interface	175
5.7	Comparison of action-types produced by supervised, heuristic and autonomous robots	183
5.8	Comparison of action styles produced by the C25K robots	184
5.9	Behaviour of the heuristic C25K robot across participants	185
5.10	Participant preferences regarding the Heuristic and IML robots	186

LIST OF FIGURES

5.11 Participant post-session evaluation of the robot across conditions 188

INTRODUCTION

Socially Assistive Robots (SARs) can be defined as robots which provide assistance through social interaction alongside or instead of physical aid (Feil-Seifer & Matarić 2005). Typical application domains for SARs include healthcare, wellbeing and education (e.g. Lara et al. (2017), Sussenbach et al. (2014), Greczek et al. (2014), Shiomi et al. (2015)). Whilst human-human interaction (HHI) is widely recognised as being critical in such domains, a lack of resources typically limits the amount of time specialist practitioners can spend with individuals. One particular example of such a resource intensive application is guiding and encouraging users through therapeutic exercises. This is an often cited application of socially assistive robotics (e.g. Malik et al. (2016), Swift-Spong et al. (2015), Wilk & Johnson (2014), Lara et al. (2017)). The motivation for such an application can be summarised as follows:

- (i) The success of therapy is related to the amount of exercise/practice the patient completes (Pollock, Gray, Culham, Durward & Langhorne 2014, Pollock, Farmer, Brady, Langhorne, Mead, Mehrholz & van Wijck 2014).
- (ii) Low engagement with such exercises is a well-documented problem (O'Shea et al. 2007, Forkan et al. 2006, Visser et al. 2014).
- (iii) HHI has been shown to positively impact on exercise uptake and adherence in a range of populations (O'Shea et al. 2007, Williams et al. 1991, Desroches et al. 2013, Jordan et al. 2010, Karmali et al. 2014) but is increasingly difficult to provide due to a lack of staff/resources.
- (iv) Preliminary studies suggest social robots can potentially have a similar positive impact on user engagement/motivation in such tasks (Swift-Spong et al. 2015, Wilk & Johnson

2014, Gockley & Mataric 2006, Lara et al. 2017); specifically moreso than an equivalent, screen-based avatar (Tapus et al. 2009).

One hypothesis to explain (iv) is that social robots go someway to emulating the equivalent human presence described in (iii), which would suggest it is reasonable to look to HHI to inform the design of effective SARs. Therapists, as experts in socially assistive interaction, are aware of the effect their presence and interactions can have. Texts on the *Therapeutic Relationship* and the *Therapeutic Use of Self* (Solman & Clouston 2016, Marilyn B. Cole MS & Valnere McLean MS 2003, Taylor et al. 2009) attempt to document best practice for using these effects to improve patient engagement with prescribed regimes. However, it is difficult to extract robot design guidelines directly from such works, predominantly because they:

- (i) focus a lot on the role of the practitioner in the context of a meaningful relationship, discussing very human capabilities and social phenomena such as seeking connection. It is not clear exactly what role an equivalent robot can or should take, and if/how the robot can or should attempt to effect this type of connection.
- (ii) do not provide sufficient low-level detail on exactly *what* practitioners should do, and *why*, that might inform the design of an equivalent robot input/action space and reasoning logic. Again discussion assumes complex human capabilities regarding social and emotional intelligence around e.g. rapport building.

As such, how to replicate this level of social intelligence on a robot, in the design and autonomous generation of such socially complex behaviour, is a key technical challenge crucial for the development and real world deployment of SARs. This work therefore aims to take an expert-informed approach to tackling this problem, looking to understand what can be learned from expert-led HHI, but also to the development and application of methodologies that allow such experts to actively contribute to robot development.

The ultimate goal of the work then will be to demonstrate meaningful, real world deployment of a fully autonomous SAR developed using this approach. Specifically, the focus will be on SARs for exercise engagement as described above, with this use case being used to set the scope and framing of the work, as further described under Section 1.3.

1.1 Related Work

Given the interdisciplinary nature of this work, each chapter details the most relevant related literature, drawing from socially assistive robotics but also social HRI, HHI psychology and other technical fields as appropriate. Here, an overview of related socially assistive robotics literature is given in order to situate the overall thesis.

Two early works in particular provide the initial motivation for improving exercise engagement as an applications of SARs. Gockley & Mataric (2006) demonstrated that, in an exercise based study, participants tended to do more exercise when they *perceived* a mobile robot to be *taking an interest* in their behaviour. Secondly, in an 8-month study looking at encouraging cognitive activities with individuals suffering from dementia, Tapus et al. (2009) demonstrated that an embodied SAR led to a more efficient, natural, and preferred interaction compared to the equivalent virtual agent on a computer screen. Together, these works suggested that a robot could have an objective impact on user engagement with an exercise task, that this was likely a social interaction/influence phenomena (linked to users' perception of *social intelligence* on the part of the robot) and that a physical robot rather than a virtual agent was likely to be more effective for this use case.

Building on this, other works have investigated the impact of particular social behaviours in SARs for exercise, with many pointing to the need for personalisation of SAR behaviour. This was first suggested by Tapus & Mataric (2008), who discussed the link between personality, empathy, physiological signals and task performance in the context of SARs. The authors pointed out that no work on SARs to date had considered the role of user profile, personality or preference in socially assistive HRI. In Tapus et al. (2008) the authors built on that notion by investigating the impact of user-robot personality matching, manipulating robot proxemics, speed/amount of movement and verbal communications to have the robot appear more introverted or extroverted. The authors also attempted to use reinforcement learning to adjust these robot 'personality' cues based on the users' task performance. The results provide first evidence of user preference for personality matching and the effectiveness of robot behaviour adaptation based on personality as well as task performance.

More recently, Swift-Spong et al. (2015) investigated the effects of comparative feedback from a SAR on self-efficacy in post-stroke rehabilitation. The results demonstrated that participants receiving other-comparative feedback from the SAR (*You had an average time that was s seconds better than others with your ability level*) might be detrimental to user performance compared to self-comparative (*During the past few trials, you averaged a time of s seconds faster than we would have predicted based on your prior performance*) or no comparative feedback. This provides a clear example of how specific implementation of social HRI behaviours can objectively impact on user performance, and hence *why* it's important that they are carefully considered. In another, very recent work, Fitter et al. (2020) again demonstrated the potential for SARs to encourage engagement with exercise, and specifically demonstrated the value of physical touch within related interactions, however their SAR implementation was not fully autonomous and did not consider the personalisation effects highlighted by Tapus & Mataric (2008).

Very few works on SARs for exercise have specifically addressed the automation of these more complex social interaction behaviours. Works on automation have instead typically focused on the generation and monitoring of sessions rather than *social* interaction behaviours. For example,

González et al. (2015) presents a planning architecture for an autonomous, NAO robot based system that can generate and take users through a custom therapy plan; monitoring sessions and demonstrating movements. However, very little attention is paid to generation of complex *social* robot behaviours, such as what encouragements to give and when. Similarly, Mead et al. (2010) presents a SAR architecture for managing task oriented interactions and generating *task performance* feedback, with no consideration of more general social behaviours. One more recent work has focused specifically on the automatic recognition of exercises necessary for SARs to be able to autonomously monitor user performance (Martinez-Martin & Cazorla 2019); whereas another demonstrated a gait training system that could autonomously monitor task performance but required teleoperation of the attached social robot companion (Leme et al. 2019).

An exception is Sussenbach et al. (2014) who implemented a motivational interaction model specifically for the purposes of autonomously generating the right type of feedback (including general, social encouragements as well as comments on task performance) to give under what conditions. The model was implemented on a NAO robot used to guide users through indoor cycling exercise. Whilst the system wasn't designed *with* an expert, it was designed based on ethnographic observations of a human fitness instructor, and so demonstrated the transfer of social interaction patterns from expert-led HHI to HRI. Another exception is the recent demonstration of a SAR whose 'emotional state' and subsequent behaviour dynamically updated based on the emotional state of user during an exercise session (Shao et al. 2019). The SAR's emotional state was then expressed through the body language of the SAR and the choice of vocabulary when ending the exercise session. Notably whilst this *shaping* of robot behaviour based on user state was impressively both *autonomous* and *dynamic* with regards to (autonomously assessed) user state, its *generation* was fixed; i.e. the same actions were always generated at the same points of the interaction.

In short, no previous works appear to have realised an end-to-end autonomous SAR that is able to provide effective, personalised, adaptive social *and* task-related feedback in the context of encouraging exercise engagement. This is of course due, in part, to the sheer complexity of generating such socially assistive interactions, whether in HRI or HHI. The initial work by Tapus et al. (2008) leveraged both psychological models of personality types *and* real-time reinforcement learning in order to inform personalisation of SAR behaviour and still concluded that resultant behaviour could be made more socially intelligent. It is posited by this work that using expert-informed approaches might offer something new in tackling this problem, a notion given credibility by the work of Sussenbach et al. (2014) described above.

Whilst expert-informed approaches have been used in the context of acceptability studies regarding SARs for exercise (e.g. Wilk & Johnson (2014), discussed in detail in Chapter 2) no studies have yet looked to work *with* experts directly to understand exactly how they utilise social interactions in order to improve engagement with their prescribed exercises. However, such approaches have been demonstrated in other SAR applications, for example, Azenkot et al. (2016)

and Lee et al. (2017) both demonstrated multi-session, participatory design processes in which experts and end-users were invited to contribute to SAR design. This work looks to take a similar approach, but also go further to consider methodologies that would also allow for expert input at the *automation* stage of development.

1.1.1 Mutual Shaping

On investigating research practices in robotics, Sabanovic (2010) found that most roboticists take a technologically deterministic view of the interaction between robotics and society. This view suggests that society should accept and adapt to robotic technologies, whose social impact is predominantly defined by their technological capabilities. However, studies in human-robot interaction (HRI) have demonstrated that the impact of robots, when deployed in the real world, is influenced by a number of social and societal factors beyond this. For example, use and acceptance of a socially assistive robot, designed for health promotion to older adults, was found to be influenced by a diverse range of factors beyond just its practical functionality (de Graaf et al. 2015). These included factors related to the context of use (e.g. social influence, privacy) and user characteristics (e.g. age, type of household). This example demonstrates the impact of real world situational factors on use of the robot; however use of the robot can also influence and change such factors in return.

As another example, the PARO robot is primarily designed to reduce stress akin to animal therapy¹. Chang & Šabanović (2015) undertook long term observations of PARO being used in a care home, using a social shaping framework to understand social factors that affected its use. Similar to the previous study, they found that use of the robot was influenced by situational factors such as the users' gender, and that robot use was often prompted by another social actor (e.g. fellow resident, staff or visitor). In addition however, they also noted that use of the robot influenced staff approaches to care. For example a carer might move certain patients towards PARO when they would normally be guided elsewhere, or use PARO as a focal point during rehabilitation therapy sessions. This shaping of robot use by the social context and shaping of the use context by robot deployment is evidence of mutual shaping on robot deployment.

However, mutual shaping is not limited to use of a robot on deployment. Sabanovic (2010) proposed the framework of mutual shaping for social robot design, with a focus on the dynamic interaction between robotics and society at all stages of design, development and evaluation. She proposed that one way to achieve this was to encourage user and stakeholder participation throughout the design and development process; representing a *mutual shaping approach* to robot design. This work looks to employ a mutual shaping approach throughout, specifically by:

1. encouraging stakeholder participation throughout design, development, automation and evaluation (employing participatory methodologies wherever appropriate)

¹<http://parorobots.com/>

2. considering the proposed SAR/SAR behaviours within its overall context of use and the societal factors that may therefore affect/be affected by its deployment
3. considering mutual shaping effects surrounding the real world deployment and use of SARs *in-the-wild*.

1.2 Research Questions

The overall goal of this thesis is to demonstrate meaningful, real world deployment of a fully autonomous, socially assistive robot, using expert-informed methodologies and a mutual shaping approach to design and development. It is posited that in doing so, the work will demonstrate that such an approach is *vital* for the successful deployment of SARs, whilst also making *generalisable* contributions to state of the art in socially assistive robotics and social human-robot interaction (HRI). Achieving this goal requires significant, interdisciplinary research work tackling a range of more specific research questions. These research questions are presented and addressed within each chapter. Overall however, in tackling this goal, the broader research questions addressed by the work are as follows:

- RQ1 How can human, domain expert knowledge, particularly regarding intuitive and experience-based social/emotional skills, be captured and utilised in the design *and* automation of a SAR?
- RQ2 Does application of expert-informed and mutual shaping approaches result in SAR behaviours that are successfully able to improve user engagement with a task/programme? Where *success* considers multiple contributory factors e.g. acceptability?
- RQ3 Considering a SAR within its broader context of use, what is the role of the human (i) designers/programmers *behind* its design/development and (ii) expert practitioners working *with* the robot ‘*in-the-wild*’?

1.3 Scope

Tackling such an ambitious goal requires a number of open research questions to be addressed. Also in order to make it achievable, a set of carefully considered constraints, limitations and assumptions need to be applied.

A huge range of robot designs and applications could fit under the very broad definition of a SAR, just as the nature of socially assistive HHI varies. This work will therefore focus on one specific application of a SAR, the previously cited example of guiding and encouraging users through a set of prescribed exercises. Specifically, the aim will be to target improved engagement with the kind of monotonous, repetitive exercise programmes that typically suffer from a lack of

adherence. This application represents a realistic, exemplar use case which allows this work to be grounded in a tangible context of use whilst also generating insights relevant to other SAR applications.

Even limited to this one application, there are a number of ways a robot might be designed to achieve the desired effect. As described under Section 5.1.2, early work on this topic demonstrated that a very non-humanoid and arguably low-sociability robot (a Pioneer mobile robot platform²) might be able to influence instantaneous task engagement (Gockley & Mataric 2006). A more recent work had the same aim, but used the much higher fidelity, humanoid NAO robot³ (Sussenbach et al. 2014).

For this work, a single existing robot platform will be used throughout, such that work will be centered on the design, testing and automation of behaviours for this platform. Whilst the overarching aim is to have experts inform decisions on exactly what role and functionality the SAR should take, the choice of platform will have a significant impact on this. Interaction modalities will obviously be somewhat limited based on the robot's hardware, but its particular embodiment is also likely to generate particular expectations around what functionalities the robot should have/role it should take. This will be recognised and managed as a limitation in the work (e.g. by introducing expert participants to other example platforms) but further study of this potential impact of robot embodiment is out of scope for this work.

1.3.1 Choice of Robot

Of existing, commercially available robot platforms, the Pepper robot⁴ was identified as being most appropriate option for the proposed use case (encouraging adults to exercise) on the following basis:

1. The size of Pepper is more appropriate for interacting with adults than smaller robots (e.g. NAO).
2. The tablet mounted on Pepper's chest can be used to present exercise instructions, video demonstrations etc.
3. Pepper's CE marking⁵ makes it easier to safely deploy the robot in-the-wild, outside of the research environment.

1.3.2 Framing

Making detailed design choices regarding the SAR role and functionalities forms a significant part of the work to be actively done with experts. However, the identified use case, choice of robot

²<https://www.generationrobots.com/media/Pioneer3DX-P3DX-RevA.pdf>

³<https://www.softbankrobotics.com/emea/en/nao>

⁴<https://www.ald.softbankrobotics.com/en/robots/pepper>

⁵<https://www.softbankrobotics.com/emea/sites/default/files/inline-files/declaration-of-conformity-pepper-1.8.pdf>

platform and overall research philosophy provide a starting point for the framing the type of SAR that is to be developed and the resulting research questions that are most of interest.

Firstly, the ultimate purpose for any developed SAR would *never* be about replacing HHI. Rather, the purpose should be to offer *additional* social interaction(s) and/or assist with that HHI in some way (e.g. by freeing up limited practitioner time for more valuable activities). As such the robot should be seen as ‘filling gaps’ in between interactions with human practitioners, and should be considered as either a tool/team member to those practitioners.

Secondly, based on the use case, the basic functionality of the SAR is expected to include being able to initiate, instruct and encourage users through a prescribed set of exercises. This represents an example of using a SAR to prompt or encourage particular user behaviours, as can be seen in other specific SAR applications (e.g. prompting children to engage in learning activities (Shiomi et al. 2015)). Clearly, to do this, the SAR will need to have some recognition of users’ performance at the prescribed exercise(s). However, detailed kinematic analysis and evaluation of said exercises are out of scope for this work. Technical implementation of such analysis would ultimately be required for full realisation of the SARs considered in this work, and must further be of high accuracy in order to be useful. Such analysis is being worked on e.g. in the context of sports analysis⁶ and weight training applications⁷. In fact, works on SAR automation to date have typically focused more on automating this type of analysis, for monitoring user performance, than on the autonomous generation of appropriate *social* behaviour (see e.g. Martinez-Martin & Cazorla (2019) discussed under Section 5.1.2). It is therefore assumed that such analysis *could* be implemented for the applications/studies presented in this work, but that this would require significant engineering effort and is less relevant to this work’s focus on designing, automating and evaluating *social interaction* for SARs.

Finally, this proposed functionality regarding the delivery of functional/task instructions, and the target population being adult users, means the robot is likely to take on a fairly *authoritative* role with regards to delivering the prescribed exercise programme. This contrasts with, for example, having the *user* represent the authority with the robot then being somehow reliant on their engagement (as demonstrated e.g. in the case of having children improve their handwriting by *teaching* a robot how to improve its handwriting (Lemaignan et al. 2016). This would also be an example of a SAR prompting engagement with a desirable behaviour through its social interactions with the user. As such, it must be recognised that the interaction behaviours developed and studies in this work represent only one role a SAR might take in attempting to influence user behaviour.

⁶Kinovea is open source video player for sports analysis that can be used to identify kinematics of movement: <https://www.kinovea.org/>

⁷‘Personal Trainer - Kaia’ App uses motion tracking on camera feed to offer e.g. squat technique corrections: <https://apps.apple.com/us/app/kaia-perfect-squat-challenge/id1393680040>

1.4 Research Philosophy

The research philosophy driving the approaches taken and research questions explored in this work is as much *practical* as it is *responsible*. In the authors' mind, the approaches taken and questions asked are necessary, from an engineering perspective, in order to build the best *technical* system that is also going to be *acceptable* to stakeholders and therefore *useful* in the real world.

For example, concerning the focus on expert-approaches, it seems *obvious* that any technical development targeting human-centred domains such as healthcare and education ought to look to capitalise on the wealth of practitioners' domain expertise. Healthcare workers, teachers and learning support staff, personal trainers etc. are all experts in the type of social intelligence that HRI researchers are attempting to replicate, at least to some extent, in social robots. However, it also seems *right* for such stakeholders to be included as much as possible during SAR design and development, as well as for such development to consider how use of SARs will impact on standard practice within these domains. Attempts to foster and understand this two-way interaction between domain experts and HRI research can be described as a *mutual shaping* approach (Sabanovic 2010) which is gaining traction (but still not the standard) within robotics research more broadly.

Given that SARs are typically designed to be used in socially complex applications, there are many stakeholders who may affect or be affected by their use on deployment. This could include teachers, parents, siblings, friends and classmates in the context of education, or carers, healthcare professionals, family and friends in the context of care; all representing stakeholders beyond the immediate user of the robot. Similarly the robot is likely to form part of a larger strategy or intervention, e.g. fitting in with a larger programme of study or forming one part of a care delivery plan. Mutual shaping effects are inevitable in such scenarios. Considered holistically they can represent an opportunity for making the best possible use of available technology. In contrast, failure to consider all stakeholders' views could lead to unpredicted negative consequences on real world deployment. Therefore, it is a key tenet of this thesis that taking a mutual shaping approach is not only responsible, but also *necessary* for success. Given that this approach is not yet the norm, posthoc critique of its implementation represents an additional contribution of this work, and is thus discussed in detail in Chapter 5.

More broadly, efforts are taken to pursue a responsible innovation approach⁸ throughout the work. This is reflected e.g. in the consideration of a recent standard for ethical robot design in Chapter 3, but also the extensive use of qualitative data collection methods alongside the more typically used categorical or Likert scale questionnaires. Such methods allow participants' greater freedom in contributing to the work, but are also demonstrated to be vital in generating deeper understanding regarding answers to/behaviour on those other measures.

⁸<https://epsrc.ukri.org/research/ourportfolio/themes/healthcaretechnologies/strategy/toolkit/home/integrity/ri/>

1.5 Structure

The chapters in this thesis essentially represent incremental, but somewhat discreet work packages undertaken in pursuing the overall goal of this doctoral research project. As such, each individual chapter introduces and addresses its own set of specific research questions/hypotheses. Similarly, a review of the most relevant literature is given within the introduction to each chapter. The broad research questions set out in this chapter are then addressed, by drawing together results and findings from across these chapters, in the Conclusion. A summary of each chapter is given below.

- Ch. 1** This chapter introduced the ultimate goal of this doctoral research work: the expert-informed design and automation of SAR, meaningfully deployed in the real world. The scope of the work was also set out, identifying the specific SAR application and framing the kind of interactions to be considered, as well as some constraints and limitations applied in order to make this goal achievable.
- Ch. 2** Chapter 2 presents an in-depth, qualitative study with therapists, as experts in socially assistive interactions, designed to understand the role of social interaction in encouraging compliance with a prescribed exercise programme. Results from five focus groups and eight interviews (with a total of 21 therapists) are collated to produce a set of (generalised) design guidelines for socially assistive HRI.
- Ch. 3** Chapter 3 identifies a model of persuasion from the HHI literature that appears to align well with results from the study with therapists in Chapter 2. A study is presented to investigate the efficacy and acceptability of SAR behaviours designed using this model. Results support this application of the model with two of the three of the tested persuasive strategies leading to participants undertaking significantly more repetitions of an open-ended exercise. It is highlighted that these behaviours may be ethically hazardous, according to a published standard on ethical robot design, and so a further two studies consider the potential impact of re-designing these behaviours to better conform to the standard. Resultant, practical design implications for SARs and social HRI more generally are also presented alongside the potential ethical considerations which ought to be made on their implementation.
- Ch. 4** Chapter 4 presents the technical implementation and real world experimental deployment of a SAR which utilises a state-of-the-art interactive machine learning (IML) system to achieve automation. This work builds on all of the design guidelines and results generated from the earlier chapters, but specifically addresses *automation* of a SAR, successfully meeting the ultimate goal of this project as set out in this chapter. Specifically, a naive ‘robot coach’ for delivering the UK National Health Service (NHS) Couch to 5km programme is co-designed and developed with a domain expert (fitness instructor) who then *teaches* the

robot how to behave in-the-wild. The robot was deployed for 11 weeks, and successfully ran 2+ autonomous sessions for each of the 9 participants by the end of the study.

Ch. 5 Chapter 5 reflects on (i) the mutual shaping approach taken and (ii) examples of mutual shaping observed throughout. Two generalisable methodologies for how to take a mutual shaping approach (based on the focus group methodology used in the study with therapists of Chapter 2 and the co-design IML approach to automation demonstrated in Chapter 5) are presented alongside results demonstrating why this is worthwhile.

Ch. 6 Chapter 6 first provides a summary of the work in each chapter and how the chapters relate to each other. The research questions presented here are then returned to and addressed by bringing together key findings from across all of the chapters. A note on ethical considerations arising from the work is given, as well as a statement regarding limitations and future work, followed by a concluding summary and list of resulting contributions.

A STUDY WITH EXPERTS IN SOCIALLY ASSISTIVE INTERACTION

Therapists are experts in socially assistive interaction, actively building relationships with service users in order to encourage compliance with/engagement in therapeutic exercises.

This chapter presents a study with therapists, designed to investigate exactly *how* they do so, in order to inform socially assistive robot design. The results highlight the importance of social influence, which therapists knowingly cultivate and leverage in order to have impact. Generalised design implications derived from the results are presented for application to a range of socially assistive robot scenarios. The work in this chapter is described in the following publication, which received Best Paper Award:

Winkle, K., Caleb-Solly, P., Turton, A. and Bremner, P., 2018, February. Social robots for engagement in rehabilitative therapies: Design implications from a study with therapists. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (pp. 289-297).

2.1 Introduction

In conventional therapy and exercise-related psychology literature, human-human interaction has been shown to positively impact on exercise uptake and adherence in a range of populations (e.g. O’Shea et al. (2007), Williams et al. (1991), Desroches et al. (2013), Jordan et al. (2010), Karmali et al. (2014)). However, it is increasingly difficult for health practitioners to provide large amounts of dedicated one-to-one support due to a lack of staff/resources. Preliminary studies suggest social robots can potentially have a similar positive impact on user engagement/motivation in therapeutic exercise tasks (e.g. Swift-Spong et al. (2015), Wilk & Johnson (2014), Gockley & Mataric (2006), Lara et al. (2017)); specifically more so than an equivalent, screen-based avatar (Tapus et al. 2009). One hypothesis to explain this phenomenon, is that social robots go

someway to emulating the equivalent human presence described in (iii), which would suggest it is reasonable to look to human-human interaction to inform the design of effective SARs.

Therapists, as domain experts in socially assistive interaction, are certainly aware of the effect their presence and interactions can have, and have attempted to document best practice for its use in improving engagement (c.f. *The Therapeutic Relationship, the Therapeutic Use of Self* as described in Solman & Clouston (2016), Marilyn B. Cole MS & Valnere McLean MS (2003), Taylor et al. (2009)). However, it is difficult to extract robot design guidelines directly from such works (which were not undertaken with the explicit aim of informing robot/technology design), predominantly for two key reasons. Firstly (i) such works often focus a lot on the role of the practitioner in the context of a *meaningful relationship* (Solman & Clouston 2016) discussing very *human* capabilities and social phenomena such as *seeking connection*. It is not clear exactly what role an equivalent robot can/should take, and if/how the robot can/should attempt to affect this type of connection. Secondly (ii) such works do not tend to provide the kind of explicit details on *what* practitioners should do and *why* that might be used to inform e.g. the design of robot actions or control heuristics.

This chapter primarily addresses (ii) using results from a study with therapists, as experts in social influence in a socially assistive domain, to improve understanding of socially assistive human-human interaction and inform SAR design. Item (i) is also somewhat addressed here with regards to what functionality a SAR might provide, but the potential for impacting on the therapist-client relationship and other considerations regarding real world deployment are discussed in the context of mutual shaping in Chapter 5. Key contributions from this work are as follows:

1. First study to take an expert-informed/user-centred approach to informing the design of SARs for therapy
2. Detailed consideration of *how* therapists use socially assistive human-human interaction to tackle service user compliance and engagement
3. Resultant design implications for socially assistive human-robot interaction

In designing this study, a novel focus group methodology was developed in order to support the overall mutual shaping approach employed throughout this research. Details of this methodology, how it compares to other participatory and user-centred design methods, and additional results and contributions demonstrating exactly how it supports this approach, are given in Chapter 5.

2.1.1 Related Work

Kang et al. (2005) undertook one of the first feasibility studies on SARs for engagement in therapy, demonstrating their potential by demonstrating a hands-off robot for encouraging breathing exercises in a hospital setting. Gockley & Mataric (2006) then demonstrated that even a very

simple, non-humanoid mobile might have an impact on compliance with stroke rehabilitation exercises. Importantly, they found that this impact appeared to correlate with participants' *perception* that the robot was somehow 'interested in' or responding to their exercise behaviour. This was further developed by Tapus & Mataric (2008) who provided initial evidence that SARs which appear to match the user in terms of extroversion/introversion might be more effective in encouraging stroke rehabilitation exercises. A more recent study considering what types of feedback SARs might give again demonstrated the utility of SARs in rehabilitation, with participants showing improved task performance gained by working with the robot (Swift-Spong et al. 2015).

Whilst these studies demonstrate the potential for SARs in therapy, they are primarily concerned with testing feasibility and quantifiable impact rather than exploring use cases, generating design recommendations or informing robot behaviour design. Studies designed to measure user acceptance have also typically been focused on evaluation of a complete/final system. For example, in the most closely related work considering SARs for rehabilitation, Wilk & Johnson (2014) utilised a robot demonstration in investigating the potential for a combined telepresence/SAR system to facilitate and encourage engagement with stroke therapy. Residents and caregivers from a daycare centre were given a demonstration of the robot's capabilities. Then, they were asked to complete a survey measuring perception and acceptability of the robot system. The authors note that caregivers also discussed additional capabilities the robot could have, but no detail is given as to the format or formality of these discussions.

Considering SARs more generally, research on robots for the care of older adults has typically employed user-centered design methods to elicit user views or assess user needs for informing design requirements (e.g. Louie et al. (2014), Wu et al. (2012), Beer et al. (2012)); but none of these works have attempted to actually understand practitioner behaviour for informing robot design. Instead, previous attempts to design social robots and other assistive technologies for engagement and motivation have typically utilised theoretical models from psychology (e.g. behaviour change theory (de Vries et al. 2017)), ethnographic observations of human interactions (Sussenbach et al. 2014) or machine learning (e.g. Chan & Nejat (2012), Leite et al. (2011)).

Outside of robotics, Singh et al. (2014) undertook a number of studies with people with chronic pain and specialist physiotherapists to identify opportunities for interactive technological devices to aid in motivation. Their aim was to consider such opportunities from two perspectives, the practical needs of those suffering chronic pain and the expert behaviour of physiotherapists in supporting them, with the latter being particularly relevant to this study. The authors used role-plays, focus groups, interviews and observations to explore physiotherapist behaviour, identifying key behaviours around the use of positive feedback, promoting self-esteem, using prompts to direct users' attention and careful choice of language. They noted that the type, and quantity of feedback provided by physiotherapists is based on the psychological state of the patient and on where s/he is in the context of their confidence with exercising. Further they noted the potential

for clinician and patient perspectives to raise potentially contradictory issues that all must be considered for any resulting technology to be effective. This provides one motivation for limiting the work here to a study with therapists (rather than patients) in the first instance.

2.1.2 Research Questions

Whilst this study is grounded in the SAR application of supporting therapy engagement, the aim is to provide insight that can be applied to a range of SAR use cases. As such, the overall research questions tackled by this study can be generalised as follows:

RQ1 How could SARs be useful in an assistive scenario, according to domain experts?

RQ2 How can user engagement with a long-term, assistive activity be measured?

RQ3 What is the role of the domain expert in influencing user engagement with the desired task/activity?

RQ4 How might SAR behaviours be tailored to individual users?

The novel focus group methodology employed in the first part of this study, designed to facilitate a mutual shaping approach, also generated valuable insight regarding mutual shaping effects that ought to be considered for real world robot deployment of SARs in therapy. This is discussed further, alongside full presentation of the focus group methodology, in Chapter 5.

2.2 Materials and Methods

5 focus groups and 8 interviews were undertaken with therapists from a range of disciplines (occupational therapy, physiotherapy, sports rehabilitation therapy and speech & language therapy) as listed in Table 2.1 (total pool N = 21, 3 male & 18 female, average age 40). Note that all interview participants apart from SL4 also took part in a focus group. Therapists were recruited by email communications to local hospitals, private practices, through advertising to university staff and through communications to contacts of the research team. Demographic information collected included time since qualified, time spent practising since qualified and typical service areas/users worked with. All focus groups and interviews were carried out at the Bristol Robotics Laboratory. The study was approved by the ethics committee of the Faculty of Environment and Technology of the University of the West of England (UWE REC REF No: FET.17.02.022).

2.2.1 Focus Groups

A novel focus group methodology was designed to facilitate mutual learning as part of the mutual shaping approach taken to all work presented in this thesis. This methodology is presented in

Study	Participants
Focus Groups (N = 20)	8 Physiotherapists (P1 - P8) 7 Occupational Therapists (OT1 - OT7) 3 Speech and Language Therapists (SL1 - SL3) 2 Sports Rehabilitation Therapists (SR1, SR2)
Interviews (N=8)	3 Physiotherapists (P1, P2, P6) 2 Occupational Therapists (OT1, OT6) 3 Speech and Language Therapists (SL1, SL2, SL3)

Table 2.1: Study with experts complete participant list.

Focus Group	Participants	Duration
1	OT1, OT2, OT3	81 minutes
2	SR1, P1, OT4, SL1	74 minutes
3	P2, P3, P4, P5	57 minutes
4	SL2, OT5, P6	76 minutes
5	SL3, P7, SR2, OT6, OT7	104 minutes

Table 2.2: Focus group participant groupings. OT = occupational therapist, P = physiotherapist, SL = speech and language therapist, SR = sports rehabilitation therapist.

full, and generalised for application to other scenarios, in Chapter 5. For this study, focus group sessions consisted of discussions, a group activity, and a project talk with demonstrations using the robot Pepper¹, lasting between approximately 60 and 100 minutes. However, as discussed further below, focus groups participants were made aware of other social robots as it was made explicitly clear that Pepper represented just one specific design/implementation of a social robot.

The ordering of these elements and key topics covered at each stage is shown in Table 2.3 and detailed below. The full topic guide is provided in Appendix A. Figure 3.2 shows the room layout employed for each session. Participants were randomly allocated to one of the sessions based on availability; group participant lists are given in Table 2.2.

Pre-Focus Group Questionnaire

A set of questions to measure acceptance of social robots in therapy were designed based on the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al. 2003). A copy of the pre-focus group questionnaire is provided in Appendix A. The pre-focus group questionnaire was completed directly before the focus group began. The questionnaire asked participants to indicate their agreement with the following statements via a 5-point Likert scale (strongly disagree to strongly agree):

1. I feel apprehensive about the use of social robots in therapy
2. Social robots are somewhat intimidating to me

¹<https://www.ald.softbankrobotics.com/en/robots/pepper>



Figure 2.1: Room layout during focus groups; note the presentation screen position on which a collage of different social robot images was displayed during discussions.

Section	Key Topics/Activities
Before Discussion	Acceptance questionnaire
Discussion Part 1	Use of social robots in therapy Conventional therapy delivery Group activity on factors affecting adherence
Project Talk & Robot Demo	Study motivations Supporting literature Project aims & objectives 2x Pepper assistance scenario demonstrations
Discussion Part 2	Revisit use of social robots in therapy Useful data collection
After Discussion	Extended acceptance questionnaire

Table 2.3: Focus group schedule showing order of activities and related topics. Note that the discussion is broken into two halves; pre and post the project talk and robot demonstrations.

2.2. MATERIALS AND METHODS

3. I think social robots might be somewhat intimidating to service users
4. I think using social robots in therapy is a good idea
5. I think using social robots would make therapy more engaging for the service user
6. A social robot would be useful in supporting a therapy programme
7. I think that use of a social robot could improve the positive outcomes/success of a therapy programme

Additionally a semantic difference question asked participants whether a social robot would be most useful when they were working with the user or when the user was working at home alone. It was hypothesised that few participants would be familiar with the concept of social robots before participating in the focus group, potentially reducing the validity of questionnaire responses. To address this, the pre-session questionnaire was attached to the following definition of social robots (based on Fong et al. (2003)):

Social robots are those that can take part in social interactions with humans. They might exhibit human-social characteristics such as expressing and perceiving emotions, making conversation, establishing/maintaining social relationships, using natural cues (e.g. gaze, gesturing) and exhibiting a personality/character. Social robots might be humanoid/resemble some human characteristics but this is not always the case. A range of social robots are shown below to demonstrate this.

Alongside this description was a collage of images showing 5 different social robots (Pepper, Buddy², MiRo³, Kismet⁴ and Nao⁵).

Pre-Demonstration Discussion

The pre-demonstration topic guide was designed to elicit participants' initial (relatively uninformed) ideas on the use of SARs in therapy. This covered ways in which SARs could be useful and more generally participants' feelings, attitudes and/or concerns about their use, and was included to collect participants unbiased opinions. Specifically, participants were asked

- *What do you think about using robots to support a therapy program?*
- *How do you think that robots might be able to do that?.*

²<http://www.bluefrogrobotics.com/en/home/>

³<http://consequentialrobotics.com/miro/>

⁴<http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html>

⁵<https://www.ald.softbankrobotics.com/en/robots/nao>

During these discussions a collage of images of 10 different robots (Pepper, Buddy, Kismet, Miro and Nao as previous, plus Bandit⁶, Molly⁷, Jibo⁸, Paro⁹, iCat¹⁰ and Pioneer¹¹ (as used in Gockley and Mataric's related work as a 'social' robot Gockley & Mataric (2006)) was displayed on the presentation screen. This is shown in Figure 3.2. Pepper was positioned alongside the moderator. Participants were able to ask questions concerning SARs, associated technologies or more about the project throughout the focus group. Pepper was positioned and set-up ahead of participants entering the focus group room; left visible and in its standard 'Autonomous Life' mode. This includes a 'breathing' motion and the redirecting of gaze and body position based on visual human tracking. Audio input and output were disabled however; such that the robot was not speaking nor reacting to speech or sounds. The collage of social robot images was referred to by as appropriate during discussions, e.g. when faced with questions concerning physical design or to probe discussion points around capabilities and applications or participant reactions to Pepper.

This part of the discussion was also used to explore related conventional elements of therapy as they are done now; similar to the approach taken by Lee et al. (2017). This focused on the importance and implementation of self-practice exercises at home, e.g. therapist prescription, user reporting and therapist monitoring of such exercises. Participants were also asked about their role in motivating or encouraging service users and how they might do that. Following this, participants were asked to complete a group activity ranking different factors which may affect user engagement. Participants were presented with a list of possible factors identified from the therapy literature, but also given blank cards to complete with any factors they additionally identified. An example result from this exercise is given in Figure 2.2. As well as generating data, this part of the discussion also established the focus group participants as 'experts'. Qualitative research guidelines (Curry 2015) suggest that expert establishment in focus groups is key to encouraging participation. Participants then feel confident that they can offer useful, valid contributions and are therefore less hesitant to take part. All participants worked with self-practice of some form, and found 'common ground' on the need to motivate service users. This helped to set all participants equal and create rapport; again important for maximum participation.

Project Presentation & Robot Demonstrations

A key aim of the study was to achieve mutual learning between the researchers and the participants; as described by Lee et al. (2017). The project presentation and robot demonstrations represent a key mechanism for participant learning and the development of mutual trust (another key factor for participatory design). They were designed to i) equip participants with a better

⁶<http://rasc.usc.edu/bandit.html>

⁷<https://www.youtube.com/watch?v=nFZ9sUbbfe8>

⁸<https://www.jibo.com/>

⁹<http://www.parorobots.com/>

¹⁰<https://www.youtube.com/watch?v=7rCqclvf12Y>

¹¹<http://www.mobilerobots.com/ResearchRobots/P3AT.aspx>

2.2. MATERIALS AND METHODS

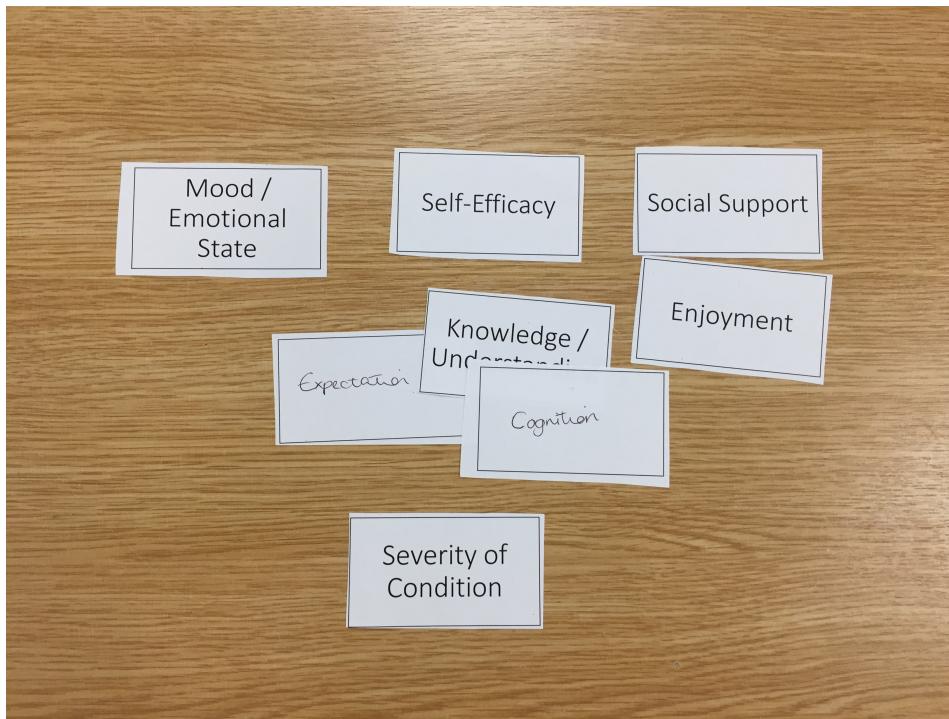


Figure 2.2: Example output of the group ranking activity undertaken in the pre-demonstration discussion of focus groups. Participants were asked to order factors which might affect user adherence to and engagement with exercises in terms of impact.

understanding of socially assistive robotics and ii) explaining the research aims and objectives, and the researcher's desire to work with therapists to design an appropriate and effective system. Following completion of the ranking exercise, participants were given a brief (approximately ten minute) presentation about the research project which covered:

1. *Background robotics literature*: was presented as evidence suggesting that robots might positively influence engagement with therapeutic exercises (specifically highlighting Gockley & Mataric (2006) and Tapus et al. (2009)). Key points included i) embodied robots seem to have greater impact than a screen based avatar equivalent and ii) a robot which seems interested in what you're doing may encourage you to work for longer.
2. *An overview of research aims, objectives and activities*: it was explained that this study was part of a larger research project exploring the use of SARs to encourage engagement in therapeutic/health-related exercise, lack of which is a known issue in rehabilitative therapies. It was explained that the ultimate aim of the project was to test a prototype with real users, and that the aim of this study was to i) generate some initial design guidelines and inspiration for that system based on the therapists' expert knowledge and ii)further the researcher's understanding of conventional therapy delivery.

3. *An introduction to the robot demonstrations:* it was made clear the demonstrations were designed just to give an initial idea of what such a robot may ‘look like’ (including e.g. in terms of its behaviour and functionality). The engineering (rather than healthcare) expertise of the researcher was noted in relation to creating the demonstrations. However, it was explained that the demonstrations were designed in conjunction with an occupational therapy researcher. Participants were also prompted to consider what they would have done as a demo themselves, what they would have liked to see and any ‘can the robot...’ questions they might have after watching the demo.

As discussed above, in the pre-demonstration discussions participants were initially asked very broadly how they thought SARs could be used in therapy. The presentation was then used as an opportunity to share a specific proposed SAR application, i.e. for motivation and engagement in exercises *in-between* conventional therapy sessions. It was highlighted that such a system was not designed to replace the therapist, but as a tool to support them. The time taken to explain *why and how* the researcher had identified that proposed application, i.e. the description of background literature described in the project presentation, was an attempt to foster mutual learning as per participatory design. Ahead of the demonstrations, participants were told that the demonstrations were very much a ‘first draft’ of possible robot behaviours designed only to illustrate the robot capabilities and give two examples of how such robots might be assistive in a therapeutic context. Afterwards, participants were also reminded that the final aim of the research project was to produce more complex, personalised and responsive robot behaviours. As stated previously, up until this point Pepper operated in autonomous life mode but with audio input turned off. To initiate the demonstrations, audio input was re-enabled via a head touch, at which point the demonstrations were launched via voice command. After the demonstrations, Pepper was shut down manually using the chest button and displayed its standard shutdown behaviour animation.

The first demonstration showed Pepper facilitating a wrist exercise taken from a leaflet on Tennis Elbow produced by Arthritis Research UK¹². Pepper explained how to do the exercise with reference to images (taken from the Arthritis Research leaflet) displayed on its tablet. It then mimed checking the user’s motion, counted repetitions for three sets of five exercises and gave encouragement. The second demonstration showed Pepper facilitating preparation of a microwave meal. Pepper gave step by step instructions and prompts, again with reference to images on the tablet. The demonstrations were live, with the researcher playing the role of a user (undertaking the requested exercise, providing verbal responses of yes/no as appropriate etc.). Pepper was operating autonomously throughout the focus group, interacting with the researcher directly.

¹²<https://www.arthritisresearchuk.org/arthritis-information/exercises-to-manage-pain/tennis-elbow-exercises.aspx>

Post-Demonstration Discussion

The presentation and demonstration made participants aware of the potential to use a SAR for motivation and engagement, and the literature that supported this approach. The post-demonstration discussion was therefore more in line with user-centered design methods, with participants being invited to give feedback and perspectives on this pre-defined research agenda. The resultant data generated in this section was informed by participants' increased understandings of SARs and their capabilities and therefore allowed for more grounded/focussed discussion. Post-demonstration discussion centred on participants' reactions to the demonstrations and a revisit of the discussion concerning the use of social robots in therapy. Additionally, participants were asked whether there was any particular data about the user that the robot could collect which would be useful for therapists to see.

Post-Focus Group Questionnaire

Directly after the focus group had been concluded and the audio recording equipment turned off, participants were asked to individually complete another questionnaire. This questionnaire contained a duplicate of the pre-session acceptance questionnaire (in order to investigate the impact of participation) plus some additional questions. These questions, designed to ensure maximum data capture and give a final opportunity for ideas and comments, were as follows:

1. In terms of achieving user behaviour change, what specific activities do you think a social robot could help with and how? (e.g. activity prompts & feedback for exercises)
2. Which types of service user group(s) do you think could benefit from the use of a social robot?
3. Any other comments?

2.2.2 Interviews

The interview schedule, given in Table 2.4, was designed to explore the role of the therapist with respect to user engagement in more detail. It was refined based on the results of the focus groups, which suggested the importance of personalised approaches for different service users. Interview discussion points and activities included reflection on two, different service users, application and evaluation of a potential categorisation tool and structured discussion. The ordering of these elements and the related discussion points are given in Table 2.4 and detailed below. The full topic guide is provided in Appendix A. Interviews were carried out on a 1:1 basis, at a later date to the focus groups, and lasted between 50 and 105 minutes.

Reflection on Two Service Users

Ahead of the interview, participants were asked to think about two service users 'who have different levels of adherence, engagement or differing motivational needs'. The interview began

Section	Key Topics/Activities
Reflection on Two Service Users	Unprompted description of pre-selected service users Reasons for choosing these users, given the research questions How participant works differently with these service users, and why
Categorisation Activity	Introduce concept of categorisation in response to focus group finding on personalisation Introduce NHS Healthy Foundations Segmentation Kit (Department of Health 2010) Application of NHS Categorisation to participant selected service users Brainstorming of custom categorisation framework or other approaches: user traits (and identification of), related approaches
Structured Discussion	Use of feedback: technical vs. motivational, positive reinforcement, triggers Progressive conditions: reflecting on progress for motivation, long-term motivation

Table 2.4: Interview schedule and topic guide.

with the participant being asked to describe them, initially without any further prompting, to investigate what traits/descriptors the participants naturally referred to. Participants were then asked what made them choose those two service users in particular, before being asked for more specific details on what factors might account for any differences between them, and how exactly the participant might work with them differently. This was recorded in real-time on flip-chart paper to ground discussions, an example output is shown in Figure 2.3. Note the later addition of NHS personas (*Balanced Compensator* and *Unconfident Fatalist*) to the service user profiles.

Application and Evaluation of a Categorisation Tool

In order to explore the concept of a categorisation framework, initially identified as one possible method of generating semi-personalised robot behaviours, the National Health Service (NHS) Healthy Foundations Life-stage Segmentation Model Toolkit (Department of Health 2010) was used to prompt discussions. This tool identifies 5 personas with different motivation to engage in a healthy lifestyle, and was designed primarily to inform health intervention design; two example personas are given in Figure 2.4. Participants were asked whether they felt the tool was applicable to their service users, or could be helpful when thinking about personalisation. This was used as a basis to prompt further listing of how and what user traits might inform personalisation of approach.

Structured Discussion Points

Based on results from the focus groups and to further inform design guidelines for robot behaviour, structured discussion was then used to investigate:

- the use of feedback: how the participant might use feedback during a session, particularly regarding motivational versus technical feedback, and what might trigger the participant to give feedback at a specific instance.
- reflecting on progress: focus group discussions suggested that reflection on service user progress could be a key motivator, but how is this dealt with in the case of progressive conditions where improvement was not expected.

2.2. MATERIALS AND METHODS

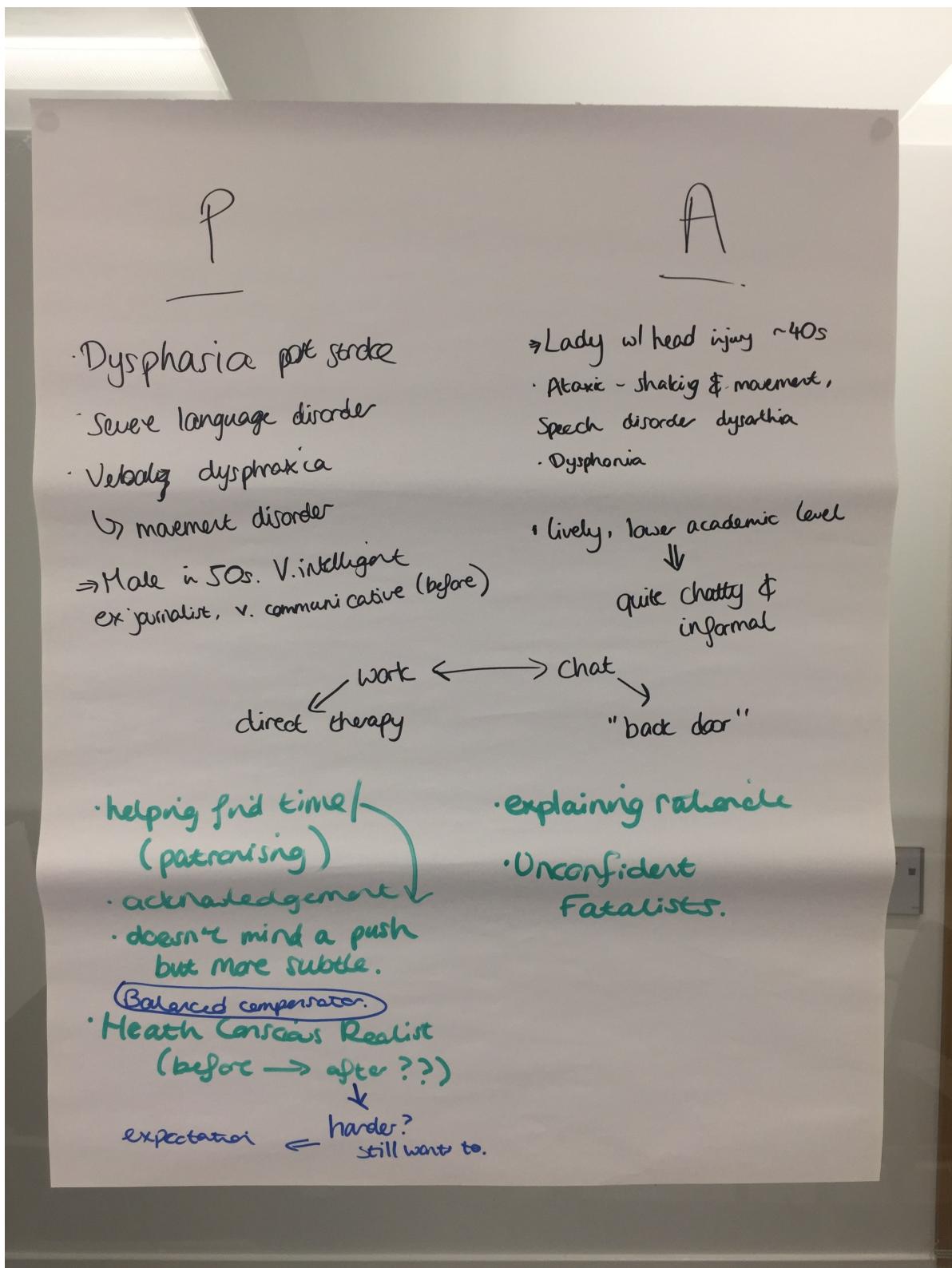


Figure 2.3: Example notes taken on participant description of two contrasting service users, as well as resultant application and identification of the pertinent NHS categorisation tool personas.

Unconfident Fatalists

- Feel negative about things & themselves
- Might be depressed
- Feel a healthy lifestyle would not be easy
- Don't feel in control of health/lifestyle
- Fatalistic, think they are more likely to get ill than others
- Lifestyle isn't healthy and their health could be better
- Know they should do something about their health
- Demotivated

◦ Older average age ◦ live in most deprived areas ◦ least likely to be in paid work ◦ more likely to be retired ◦ fatalistic about health ◦ hold negative perceptions of healthy lifestyle



Balanced Compensators (BC)

- Positive, like to look & feel good about themselves
- Understand that their actions now impact on current & future health
- Health is very important to them & something they feel in control of
- Feel a healthy lifestyle is generally easy & enjoyable
- Will compensate for health risks e.g. going for a run the morning after eating a big meal/drinking too much

◦ Male bias ◦ highest proportion of people in full time work ◦ exercise regularly ◦ eat healthily ◦ low drug/smoking use



Figure 2.4: Two different service user personas identified by the NHS Healthy Foundations Life-stage Segmentation Model Toolkit, designed to inform health intervention design.

2.2.3 Data Analysis

Transcripts of the focus group and interview discussions were (separately) analysed using the Framework method, following published guidelines on the analysis of qualitative research data multi-disciplinary health research (Gale et al. 2013). The data was first coded for key themes using a combined deductive and inductive approach to coding. In both cases, initial codes were generated based on the literature review and research questions. Two initial transcripts were then coded by the researcher as well as a member of the supervisory team, with any additional codes generated inductively as required. The resulting, individually generated codes were then compared (considering overlap, frequency of occurrence etc.) and a final coding scheme was generated for application to all transcripts.

A second level of inductive coding was then applied to specific nodes as required. The full coding scheme for both focus group and interview data (including all second and third level inductive codes) is given in Appendix A. Further according to the Framework method, themes were then identified across the data by reviewing the coded data excerpts and making connections within and between participants and coding nodes.

2.3 Findings

2.3.1 Use of SARs in Therapy (RQ1)

Initially, in the pre-demo discussions of focus groups, participants struggled to see how SARs could be useful in therapy. Some of the physiotherapists had seen and used physically assistive rehabilitation systems and commented they had been expecting to see something similar. Most suggestions referred to current technologies such as smartphone applications, computer software and fitness trackers. However, as the discussion progressed and participants were asked to reflect on factors affecting engagement, without researcher reference to SARs, many then brought up the idea of using a SAR to aid with that:

[P2]: *"If you've got a buzz on your wrist telling you to move...it does get them thinking about it, so if you've got some humanoid or puppy doing a similar sort of thing, because compliance is a massive issue"*

[P6]: *"I could really see a place for that bit between sessions to help maintain motivation and erm engagement in carrying out what otherwise could be quite mundane therapy programmes"*
(P6)

After the demonstrations participants specifically identified and commented on the ways in which a SAR might add value to existing technologies. Most of this discussion centered on the value of embodied interaction for engagement and enjoyment. 8/21 participants also named specific smartphone applications or computer software they currently use for providing feedback, that they would like to see integrated with a SAR for motivation.

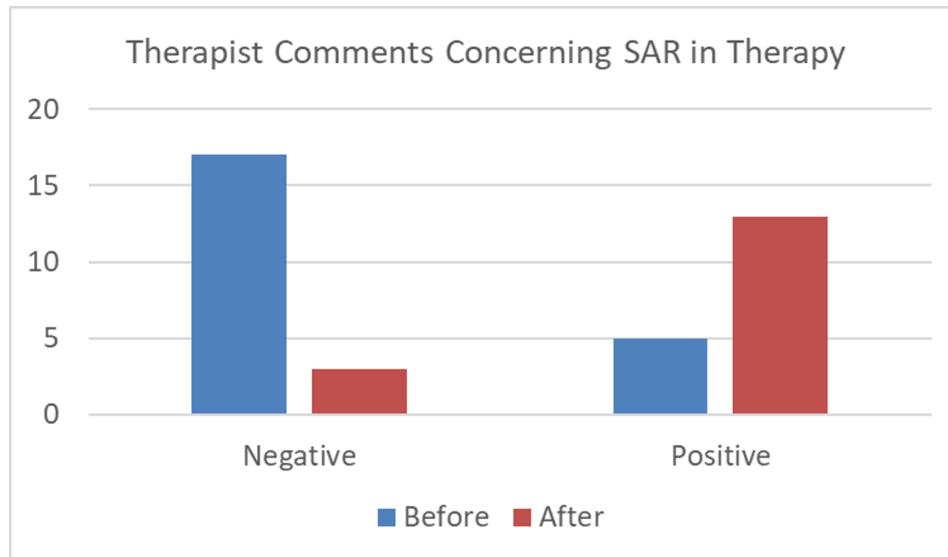


Figure 2.5: Shift in valence of participant comments before and after witnessing focus group demonstrations and project talk.

[P2]: “It’s just amazing how it moves, just that interaction with the eyes and the eyelashes and stuff so it kind of makes it feel like there’s something, a more personal feel to it or, I know personally for me, that I would work better with that than just looking at a tablet.”

[SL2]: “people with a brain injury...they will use something like an iPad and it will be like step by step instructions...they’ve got the photo and they’ve got a voice prompt...I can set it for a certain time for the alarm to go off so it’ll say ‘don’t forget you need to do x’ but I think that with this, such presence...they may be less likely to forget that they were doing something. I like that.”

[P4]: “A lot of the time we use kind of targets or something for service users to aim at and I wonder if that could be included within the screen or being able to reach out for Pepper’s hand”

These results also suggest that taking part in the focus group had a significant impact on participant acceptance. Figure 2.5 shows the shift in valence of comments coded under ‘Therapist Opinions’ and labelled as positive or negative during our data analysis. This represents an additional potential benefit of pursuing a mutual shaping approach to HRI design, discussed further in Chapter 5.

In terms of specific applications and use cases, participants immediately identified using the robot as a mediator within session when working with children. For adults, participants saw it more as a tool for service users to use in-between therapy sessions. Participants unanimously agreed that children would love to work with the robot; and many felt the benefits could be equally enjoyed by adults too.

[OT5]: “I think the value of novelty and fun is so therapeutic for adults as well...the effect, just seeing it interactive made me feel a bit happier...it’s just nice”

Participants did suggest however that there may be some service users who had simply

no interest in working with a robot. Some participants suggested this might be linked to age but others disagreed. Participants all agreed however that personalisation of the robot and its motivational strategies would be particularly important when working with adults, both for acceptance and for maximising impact.

Concerns

Initially, participants' concerns were centered on the value a robot could add to existing technologies and whether that justified any additional associated cost. This significantly reduced after the robot demonstrations, however some participants still questioned exactly how '*fancy*' the robot needed to be in order to have a positive impact. Another key concern raised by multiple participants was whether a robot could ever really be adaptive enough, or be able to '*read*' the user as discussed in Section 2.3.2.

[OT4]: "*We get our input from what we observe we don't often have the person say things we just observe them and know what we need to do, so how does a robot then observe a person without you having to instruct it?*"

More practical concerns focused on the medical needs of specific users, e.g. being able to recognise the speech of a dementia user. As these are very specific to particular use cases they are not considered in detail here, but such practical requirements must be well considered when designing assistive robots for real world use.

In terms of facilitating exercise sessions, one key concern was around giving accurate demonstrations and feedback. Participants were wary of the robot trying to technically evaluate users performance and hence suggested asking users to self-rate instead. In fact, this was highlighted as a positive thing that therapists themselves often do, in order that users learn to recognise good performance. Additionally, whilst participants liked the idea of the robot demonstrating exercises itself, they were wary of it not being able to do so accurately and so suggested it might be better to use a tablet or other conventional method.

2.3.2 Measuring Engagement with Therapy (RQ2)

Completion of self-practice and user progress were typically cited as easy to obtain long-term measures of engagement; however it was noted that user reporting of self-practice is not always accurate. Participants found it slightly more difficult to identify measures of engagement within session, but typically listed social cues such as body position, facial expression, eye contact and amount of questioning or discussion.

Significantly more discussion was focused on the recurrent theme of participants '*getting a feel*' for the user; which all participants found very hard to verbalise and explain. Typically this would be done in initial interviews or discussions, from which participants felt they could generally predict how engaged a user would be and what sort of approach might be appropriate with them.

[P1]: “*I think the really important bit is the initial subjective interview with the user, you need to know the whole psychosocial background really and understand where they are and really by identifying their goals and things you get a good idea, you get a good feel, from verbal and non-verbal communication*”

All participants highlighted how much this impacted on their approach to working with the user. Key identified user traits are discussed in detail under Section 2.3.4. Most participants further suggested that ‘reading’ a user this way was an intuitive skill, built up with experience over time.

2.3.3 Therapists’ Role in user Engagement (RQ3)

Discussions on this topic typically centered around two key themes, i) identifying and tackling barriers to engagement and ii) improving intrinsic motivation. All participants recognised their role in motivating the user.

“*Do you find yourselves having to motivate service users?*” [strong agreement from all] [P4]: “*And that’s hard when you’re working in an NHS field, you know that they need it, you know that’s their personality but you just can’t, can’t give that*”

Some participants also felt they had a role in providing external motivation to a user through generating some sense of ‘accountability’, but pointed out the aim was always for improved intrinsic motivation. This also fed into ideas around positive reinforcement.

[P5]: “*They want to be held accountable to somebody* [agreement from all] *they’re not being held accountable to themselves so they put you into a position of accountability and therefore they’re doing it to please you when in fact it should be about pleasing themselves.*”

In cases where service users were already intrinsically motivated, participants still identified the need to facilitate engagement, e.g. by helping them schedule a convenient time or suggesting prompting and data recording methods to target memory issues.

Results from the ranking exercise undertaken in focus groups made it clear that factors affecting engagement are very individual to each user. No group generated the same hierarchical ranking as another, one group concluded it was impossible and in all groups the task prompted significant deliberation, discussion and disagreement. However, some key themes concerning how therapists typically target user engagement did emerge from this task as well as additional interview discussions. These were:

- providing meaningful positive reinforcement
- improving knowledge and understanding of the user’s condition, therapy and its benefits
- personalisation of exercises based on interests
- personalisation of approach based on knowledge of user
- reflection on goals/ functional benefits of exercises

- empowerment of the user
- relationship/rapport between the therapist and user

It also became clear that, as with factors affecting engagement, specific instances of these were different for each therapist-user pair. This is discussed further in Section 2.3.4.

The Importance of Social Interaction and Influence

The list of ways that therapists target engagement presented above are almost all concerned with social interaction between themselves and the service user. As such, their effectiveness is completely dependent on the strength of their relationship with the service user, or more specifically the *influence* the therapist has over them, which determines to what extent the e.g. encouragements, prompts, admonishments they give are likely to have an impact on the service user. This is quite well documented in the literature (c.f. *The Therapeutic Relationship, the Therapeutic Use of Self* as described in Solman & Clouston (2016), Marilyn B. Cole MS & Valnere McLean MS (2003), Taylor et al. (2009)) but was also identified by study participants as something they are not only aware of, but actively work to maximise, personalising and tailoring their approach where necessary, in order to get positive outcomes.

For example, Figure 2.3 shows notes taken about two different service users as presented by participant SL1 during his interview. When identifying how he approached working with service users P and A differently, he spoke very much around the difference in how he utilised social interaction and the type of relationship he aimed to cultivate with them, even though the end-goal was the same - to improve their compliance and engagement with his prescribed exercises in-between sessions. As noted in Figure 2.3, with P he identified the need to take a direct, work-like approach whereas with A he referred to the need for much more social interaction, as per the following interview extract:

[SL1]: *A's probably informal, therapy through the backdoor and P is more direct therapy because that's what he expects. And would be disappointed if it weren't.*

[KW]: Would you consider A quite de-motivated at times?

[SL1]: *Yes I would...she needs constant support.*

[KW]: How do you do that then, is it a bit of coercion?

[SL1]: *It's a coercion yes, you can joke along with her you can have fun with her, very much make her laugh, we tend to have a really good chat, banter first of all then serious stuff and then back to the banter again.*

[KW]: A sandwich of humour and seriousness?

[SL1]: *Yes (laughs) and she would go along with this, she'd let me in the room she'd talk to me but I don't feel she would engage so much if you didn't do that around the therapy, it has to be an enjoyable experience for her I think. She has to enjoy the visit and she knows that a bit of it is going to be me pestering her a little bit about what she's got to do, she doesn't mind that if its amongst all this other stuff as well.*

[KW]: And is that true for P as well?

[SL1]: *No I think P obviously gets enjoyment out of it and enjoys the session but wouldn't necessarily need that, wouldn't need me to, we do chat alongside and we do sometimes laugh about things but it's not as important to P, I could certainly do sessions without that with him.*

2.3.4 Personalised Approaches (RQ4)

Results from the focus groups made it clear that personalised approaches are key to motivation and engagement in therapy. Taking a ‘client centered approach’ and tailoring therapy to the user more generally was raised as best practice across professions.

[P1] :*I would always take a different approach with every user, so even if I was looking at two service users with sore knees I'd be taking different approaches with those two people based on their beliefs and expectations of the treatment.”*

This was explored further in the interviews using an NHS categorisation framework for health behaviours, as discussed in Section 2.2.2. All participants could see the worth of trying to tailor behaviour of a SAR based on the user, and agreed that there was value in identifying user traits to inform that. Further, all could identify somewhat with the NHS framework descriptors and personas. 7/8 participants however struggled with the concept of labelling people into discrete categories, and preferred to instead talk about specific user traits and how those informed their approach. The user traits consistently linked with informing therapist approaches are listed in Table 2.5. Complimentary therapist approaches identified as being adaptable to each user are listed in Table 2.6.

Other factors were identified as being important to engagement and therefore would be targeted by the therapist (e.g. enjoyment of session) however these were less linked directly to specific user traits. Section 2.3.2 discussed the concept of therapists *getting a feel* for service users and their likely engagement. Many of these user traits might be identified this way, e.g. through an initial interview. Other methods of learning about a user included reading their case notes or assessing medical history, talking to friends, family or other health professionals, and observing a user’s surroundings (particularly in the home). Other factors identified by participants as relevant here, but less referred to when discussing specific approaches, include fear of exercise or anxiety, mental health, realism/ acceptance, general positivity and perceived difficulty.

2.4 Design Implications for Socially Assistive Robots

Whilst the results presented specifically consider SARs for therapy, the resultant design implications can be generalised for other assistive scenarios in which the role of the SAR is to improve engagement in self-directed activities, in conjunction with a domain expert practitioner.

Firstly, the results give support to the hypothesis that social robots can be useful, assistive companions, and that the embodiment of a SAR might help to tackle the issue of low engagement

2.4. DESIGN IMPLICATIONS FOR SOCIALLY ASSISTIVE ROBOTS

Previous activity levels/engagement in sport
Indicates whether user is likely to understand the concept of training, and what expectations they may have about returning to a certain level of activity. Informs e.g. approach to exercise programme.
Employment status
Indicates the time pressures service users are likely to be under; but may also give an idea of work ethic, education level and socio-economic situation. Informs e.g. approach to scheduling sessions.
Motivation/ Self-efficacy
Indicates how willing the user is to change intrinsically already. Informs e.g. amount of positive reinforcement given and style of motivational messages.
Cognition
Establishes how much a user can understand and remember. Informs e.g. communication style and approach to exercise programme.
Education/ Intelligence
Indicates how much user is likely to know about their condition as well as therapy and its benefits. Also raised as being potentially linked to motivation and likely linked to socio-economic situation. Informs e.g. how knowledge and information is delivered.
Family/ Social Support Situation
Indicates whether additional external motivation and support is likely to be provided.
Functional Goal(s) and/or Interests
Functional goals establish service users' main reason(s) for being in therapy, potentially giving insight into motivation levels. Interests give the therapist something to incorporate into exercise design or social interaction for rapport building. Both can be used for improving motivation.

Table 2.5: Key user traits identified as informing therapist approaches to facilitating and encouraging user engagement with therapy.

in a way in which things like smartphone applications, computer software and fitness trackers cannot. Participants felt that patients would find a SAR more engaging, more enjoyable to work with and harder to ignore, dismiss or forget about than current methods. Many other benefits of SARs were proposed by the participants, however most of these follow on from the benefits of other modern methods, e.g. a potential lack of embarrassment exercising with a device rather than a person. Further, a SAR must still offer all of the functional capabilities of existing technologies in order to be useful in the real world. To this end, the linking of SARs with existing technologies such as smartphone applications and computer software, particularly those that already deal with providing specialist technical/task-specific feedback, should be explored.

The results suggest that a SAR might help improve engagement by i) improving ease of access and ii) providing motivational support. Proposed robot behaviours and functionalities to address these aims are presented in Sections 2.4.1 and 2.4.2 respectively. Many existing methods already address these aims to some degree; so all functionalities are listed for completeness and only embellished in cases where use of a SAR specifically might offer additional value. In addition, the results suggest that SAR interaction and engagement strategies must be personalised and

Improving Knowledge & Understanding
e.g. targeted use of evidence, choice of language and level of detail when trying to help service users understand why what they're doing is important and beneficial
Exercise Style
e.g. whether exercises sessions are based more on traditional workouts or instead 'disguised', incorporated into daily activities etc.
Engagement Strategies
i.e. ways in which therapist addresses low engagement e.g. gamification/ competitiveness, distraction techniques.
Use of Feedback & Positive Reinforcement
e.g. style (challenging, technical or functional, reassuring), level of detail, amount etc.
Provision of Additional Support
e.g. exploring other health issues, providing additional lifestyle guidance such as eating/drinking prompts, relaxation techniques etc.
Incorporation of Functional Goal(s)/ Interests
e.g. reflecting on functional rather than medical progress, using interests for distraction or enjoyment.
Reminders & Prompts
e.g. whether jointly agreeing a time, an external prompt or reminder system, a method of self-reporting etc.

Table 2.6: Key elements of therapy that therapists will personalise to the patient in order to facilitate and encourage engagement.

adaptable to the user; this is discussed in detail in Section 2.4.3.

2.4.1 Improving Ease of Access

The following key robot functionalities are identified for improving ease of access to self-directed activities. Additional detail is provided in instances where SARs might offer additional value compared to current methods, as indicated by *.

1. Scheduling, reminders and prompts*
2. Facilitating the activity*
3. Data recording
4. Communication link to therapist

Scheduling, Reminders & Prompts

Ideally, the robot should approach and prompt the user to start an activity at the pre-agreed time; potentially also giving the user an advance reminder (e.g. 20 minutes ahead of time) that they are scheduled to begin shortly. How the robot deals with non-compliance should be decided as

part of a higher level, personalised engagement strategy, as discussed in Section 2.4.3. However, there is an opportunity for SARs to add value here by:

- i) dealing with non-compliance in a socially appropriate way based on the user (i.e. reacting to real-time social cues and/or as per domain expert strategies)
- ii) learning about the user to inform which behaviours to perform when, e.g. learning which social cues are linked to engagement and likelihood of compliance
- iii) adapting both scheduling practicalities (e.g. suggesting new times, reminder settings) and social behaviour in order to maximise likelihood of compliance.

Facilitating the Activity

The robot should guide the user through activities defined and set by the domain expert; ideally by displaying and referring to a video demonstration and/or demonstrating the movements itself if possible. Specifically with Pepper, it was suggested that a video demonstration would be best, as the robot would not be able to demonstrate movements in a realistic way and hence may risk having users attempt an incorrect/unsafe movement. Additional value could be generated by utilising the embodiment of the SAR during activities; e.g. using the robot end effectors as targets for exercise. Motivational impact associated with SAR interactions within session are discussed in the following session.

2.4.2 Improving Motivation

The following list represents general strategies a SAR might employ in order to motivate the user. All of these strategies are employed in some way by existing methods therapists might use, either directly or through other technologies like fitness trackers, to improve motivation. However, participants felt that embodied interaction with a social robot would add value to items 1-5 even if there was no specific additional functionality gain. Participants were also able to suggest robot specific behaviours for enjoyable interactions. These were mainly based either on physical contact between the user and the robot (e.g. a high-five) or through the robot being ‘entertaining’ through use of its body (e.g. dancing, cheer-leading etc.).

Existing literature reinforces these ideas. Previous research has demonstrated the value of embodiment in similar scenarios (e.g. Tapus et al. (2009), Wainer et al. (2007)). Furthermore, social HRI studies have demonstrated the impact of specific robot behaviours (including touch) on motivation (Nakagawa et al. 2011) and persuasion (Chidambaram et al. 2012). Based on this, it seems worthwhile to further develop and test specific, physical social robot behaviours targeting motivation as might be appropriate to specific application contexts.

As identified in Section 2.3.4, the way in which a practitioners might try to improve a users motivation is very specific to that individual; the general strategies here require personalisation to be effective. This is discussed in detail under Section 2.4.3.

1. Reflecting on activity progress & goals

2. Improve knowledge & understanding of why the activity is important
3. Provide positive reinforcement & motivational feedback
4. Allow practitioner to access activity data, and remind users of this
5. Make the activity (more) enjoyable

2.4.3 Personalisation & Adaptability

Findings from this study suggest it is vital for the robot to be personalised to the user in its overall functional, motivational and interaction strategies as well as able to adapt in real-time to user behaviour during interactions. These are hereafter referred to as *high level personalisation* and *real-time adaptation* respectively.

High Level Personalisation

The results presented in Section 2.3.4 suggest that a discrete categorisation framework of user ‘motivation type’ and associated robot settings is unlikely to achieve the necessary level of personalisation. Instead, high level personalisation will likely require the adjustment of a number of key robot characteristics based on particular user traits. An initial list of robot settings, based on the therapist behaviours listed in 2.6 is presented below. A brief description of each setting is given, along with user traits which may be relevant (based on the service user traits listed in Table 2.5). Associated personality traits have been linked to settings according to equivalent mappings identified by the therapists, as discussed in Section 2.3.4. These have still been generalised and should be applicable to other SAR scenarios, but will definitely require further refinement for other applications. Additional further refinement of traits, and testing and development with end users is required for refinement and validation in any use case scenario. To this end, *User Preference* has been included as an additional ‘user trait’ for settings which might be chosen directly by the user rather than as a result of their traits.

1. Style of Approach

Whether the robot should take e.g. a more direct, activity focused style or a more friendly, indirect approach to the required task(s). This would impact e.g. use of language, level of formality and amount of unrelated interaction.

Previous level of activity; Self-efficacy; Education

2. Reminder Protocol

The process by which the robot reminds users about/prompts users to start an activity, e.g. is it at a set time, do they get advance reminders, flexibility (e.g. how often can they say no), the way in which negative responses are dealt with etc.

Employment; Social support

2.4. DESIGN IMPLICATIONS FOR SOCIALLY ASSISTIVE ROBOTS

3. Knowledge Delivery Approach

How the robot delivers information aimed to increase knowledge of surrounding the activity and its importance. To include e.g. whether knowledge should be more technical or functional based, how often it should be delivered, method of delivery etc.

Previous level of activity; Self-efficacy; Cognition; Education; Functional goals/interests

4. Use/Desired Impact of Feedback

Tailoring of motivational messages and feedback e.g. to be more challenge focused/competitive versus reassurance based.

Previous level of activity; Self-efficacy; Cognition; Education; Functional goals/interests

5. Engagement Strategies

Strategies combating a lack of engagement during activity sessions. To include e.g. distraction techniques, gamification or increased social interaction. Further design, testing and validation of such strategies is required.

Self-Efficacy; Cognition; Social support; Functional goals/interests; User preference

6. Robot ‘Persona’

Characteristics of the robot’s persona, to include e.g. gender, voice options.

User preference

Real-time Adaptation

Building on the the idea of overall personalisation, the robot should also adapt its behaviour in real-time based on the user during activity sessions and other interactions. This adaptation should be informed by the high level personalisation described above. Whilst our study highlighted the importance of such adaptation, it was more difficult to isolate specific user cues and the therapist behaviours to which these would be linked.

The use and tailoring of feedback was one therapist behaviour consistently linked with real-time adaptation and user cues. Specifically, participants discussed being empathetic as they encouraged the patient (i.e. by recognising and acknowledging fatigue or discomfort) and using personalised feedback or positive reinforcement (as per the high level personalisation) when most needed. This correlates with a previous finding that users exercise for longer with SARs which include acknowledgement in their motivational model (Schneider et al. 2017).

In order to maintain engagement during an activity session, the robot should employ pre-defined strategies as discussed in the previous section (e.g. gamification/competitiveness, additional social interaction etc.). However, to do this it must be able to recognise when the user is disengaged. Some initial real-time measures of engagement are identified in Section 2.3.2. Existing research on automatic analysis of engagement demonstrates it is possible, but heavily dependent on user personality; further reinforcing the importance of personalisation in this

context (Salam et al. 2017). It is expected that investigating this more will require ethnographic observation and potentially coding of practitioner behaviour and practitioner-user interactions.

2.4.4 Socially Assistive Robots Need Social Influence

Based on the results from the study with therapists, it is a key postulate of this thesis that social assistance is achieved through social influence, i.e. that user behaviour is changed (as desired) by manipulation of the social environment. Therefore, SARs must also be able to impact on and manipulate the social environment in order to effect the same type of social influence. For example, the concept of user accountability to the practitioner is discussed in Section 2.3.3. Participants were unable to agree whether the presence of the robot, and it ‘watching’ and monitoring the user’s engagement would be able to replicate that phenomena. This raises the question of whose social presence a SAR would really leverage - its own, that of the relevant practitioner, or potentially both? It would be worthwhile to test user compliance with a SAR with and without practitioner data sharing capabilities in order to investigate this.

Whilst the design implications here regarding e.g. certain functionalities and the personalisation of behaviours are aligned to this requirement, further work is required to understand exactly to what extent the human phenomena of social influence can be replicated in HRI. The presence of a robot has been demonstrated to change people’s behaviour in terms of e.g. honesty and obedience (Forlizzi et al. 2004) and decision-making (Stanton & Stevens 2014); therefore it is reasonable to expect some change in user behaviour just from introducing a robot into their self-directed activities. However, few previous works in HRI have drawn from human-human social influence/persuasion literature directly, and therefore it is still to be investigated whether e.g. human models of influence and persuasion can be of use in robot design. This position also raises a number of ethics and acceptability questions which must be addressed both to ensure responsible design/development but also to ensure any resultant SARs are acceptable to, and hence utilised by, end users. These considerations are discussed in detail in Chapter 3.

2.5 Conclusion

This chapter presents findings from a study with therapists, consisting of 5 focus groups and 8 interviews (total pool N = 21 occupational, sports rehabilitation, speech and language and physiotherapists). The study was designed to investigate participants behaviour participants’ behaviour in supporting service users to engage with self-directed therapeutic exercises (as an exemplar in assistive human-human interaction), and their perception and ideas of how SARs could be helpful in this context. Key findings from the study can be summarised as follows:

- Participants identified the potential for SARs to improve engagement with therapeutic exercises, noting the impact of embodiment compared to e.g. a smartphone application

- Whilst participants identified some common measures of long-term engagement with prescribed exercises, they found it harder to explicitly identify within-session measures of task engagement, and also noted the importance of their ability to *get a feel* for a service user and their engagement
- Therapist's social influence on their service users is crucial to their ability to then provide assistive interactions - participants identified the importance of their relationship and social interaction with the service user, which they cultivated and utilised both to provide external motivation (e.g. a source of accountability) but also in then trying to improve intrinsic motivation (e.g. by discussing importance and benefits of engagement)
- Participants stressed the importance of both 'high-level' personalisation and real-time adaptation when working with service users

These findings were analysed to generate a set of design implications for informing assistive, social HRI and SAR design more generally. These include key robot functionalities, considering how SARs might add value to existing methods, and discussion of the need and type of personalisation. To this end, an initial list of key robot characteristics which should be personalised, and specific user traits these might link to, is also presented.

Arguably however, the most crucial conclusion from this study is about (i) the extent to which social influence can account for user adherence to/engagement with a prescribed exercise regime, even though such regimes have intrinsic benefits/might seem to be of obvious importance for service users' health/wellbeing and (ii) how this informs practitioners' behaviour. Practitioners are acutely aware of the role social influence plays shaping user behaviour, and explicitly aim to cultivate such effects through intelligent and purposeful tailoring of their own social behaviour and interactions with the service users. The resulting axiom for SARs can be summarised as follows:

Social influence is defined as the way in which individuals change their behaviour in response to a social environment (Gass 2015). In socially assistive human-human interaction scenarios such as healthcare, social influence is a key mechanism by which practitioners actively attempt to impact on service user compliance and engagement with beneficial activities. If socially assistive robots are to provide a similar function in such scenarios, they must similarly be able to effect user behaviour changes through social influence. The effectiveness of such robots will therefore depend on their ability to intelligently leverage social interactions and tailor their social behaviour for maximum impact on the user.

This represents a new way of conceptualising socially assistive HRI, as an attempt to shape user behaviour through social influence, novelly posited by this thesis. Such conceptualisation raises the question of whether, building on the design guidelines presented here, models of social influence from human psychology might also be used to inform SAR design/automation. This

question motivates the work presented in Chapter 2: *Persuasion as a Model for Socially Assistive HRI*.

CHAPTER



PERSUASION AS A MODEL FOR SOCIALLY ASSISTIVE HRI

The study with therapists presented in Chapter 2 led to the notion that socially assistive human-robot interaction (HRI) might be framed as the effecting of user behaviour change through social influence. This chapter firstly presents a review of relevant human-human interaction (HHI) and social HRI literature to further demonstrate how this framing is evidenced by existing literature. Three studies are then presented to (i) demonstrate the effectiveness of socially persuasive behaviours, in the context of a socially assistive robot (SAR), and (ii) consider whether such behaviours can be designed to conform with a published standard for the ethical design of robots whilst still being effective. Resultant, practical design implications for SARs and social HRI more generally are also presented alongside the potential ethical considerations which ought to be made on their implementation. Part of the work presented in this chapter (specifically limited to Study 1) is described in the following publication:

Winkle, K., Lemaignan, S., Caleb-Solly, P., Leonards, U., Turton, A. and Bremner, P., 2019, March. Effective persuasion strategies for socially assistive robots. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 277-285). IEEE.

3.1 Introduction

A key conclusion from Chapter 2 was that successful socially assistive HHI should result in desirable user behaviour change through social influence. This chapter therefore considers whether literature on social influence could usefully inform the design of SARs. Specifically, persuasion is identified as a form of influence by which one agent might effect behaviour change in another, that may be an appropriate way to model the type of task-focused socially assistive HRI considered in this work. It is posited that this modelling works at two levels:

- (i) In many applications of SAR specifically, the role of the robot is to prompt/encourage particular user behaviour(s) e.g. in the context of facilitating and encouraging exercise (Schneider et al. 2017, Lara et al. 2017, Swift-Spong et al. 2015). These can be considered instances of persuasion.
- (ii) Human persuaders intelligently engage in social/socially persuasive behaviours in an attempt to boost their (perceived) credibility and likeability. This is analogous to approaches in social robotics *more generally* that consider how robot behaviour/design can impact how a robot is perceived (e.g. Lucas et al. (2018), You & Robert Jr. (2018), Martelaro et al. (2016)) typically in an attempt to boost the ascription of desirable traits which show overlap with the human constructs of credibility and likeability (Gass & Seiter 2015).

This chapter presents a review of existing literature as well as three new studies to support this assertion, whilst also considering the resulting acceptability and ethical implications.

Study 1: ELM for Socially Assistive Robotics

A between-subject, laboratory based study designed to investigate the impact of persuasive social behaviour (demonstrations of goodwill, similarity and expertise), derived from the Elaboration Likelihood Model (a human model of persuasion), on perception and persuasiveness of a socially assistive robot.

Study 2: Varying the Source of Expertise

A within-subject, online, video-based study designed to investigate the impact of varying the source of expertise the robot refers to (i.e. referring to *its own* expertise versus the patient's *therapist's* expertise) on perception of a socially assistive robot.

Study 3: (More) Ethical Design of Social Dialogue

A within-subject, online, video-based study designed to investigate the impact of designing socially persuasive dialogue in a way that better complies with recommendations from BS 8611 (by being less anthropomorphic) on perception of a socially assistive robot.

Key contributions from this work are as follows:

1. Novel consideration of persuasion as one way to model socially assistive HRI, also (re-)framing approaches in social HRI more generally in the context of building credibility/likeability as per persuasion in HHI.
2. An HRI study in support of the above, demonstrating:
 - an exemplar SAR use case scenario showcasing the persuasive nature of SAR functionality.

- the effectiveness of HHI-inspired persuasive strategies as applied to robot design in this context.
3. Consideration of the ethical implications of designing socially persuasive robots, including two preliminary online studies investigating any potential trade-off between ethical design and effectiveness.
 4. Detailed qualitative data collection and analysis across all of the above, providing a significant insight into the complex reasoning underlying how social robot behaviours are perceived, the way they might impact on user behaviour and whether or not this can be measured through the Likert and semantic difference style questionnaires typically employed in HRI studies.

3.1.1 Persuasion and Social Influence in HHI

Social influence can result from a number of different psychological phenomena. For example, Kelman defined three varieties of attitude change, *compliance*, *identification* and *internalisation*, that might result from social influence (Kelman 1958). Other types of social influence that might result in behaviour change include *conformity* (Aronson et al. 2010), *reactance* (Brehm 1966), *obedience* (Milgram 1974) and *persuasion* (Gass & Seiter 2015). Of these, only compliance, obedience and persuasion specifically relate to one agent making specific requests of another, and hence are viable for modelling the kind of task-focused, instructor-led socially assistive interactions considered in this work.

Considering the study with therapists in Chapter 2, therapists should represent figures of authority to their service users, and obedience may therefore form part of the reason they can influence compliance with a task. However, attempts to generate obedience alone would not account for the large range of social interaction reported by the therapists. Similarly, the main difference between *compliance* and *persuasion* is that compliance refers specifically to a change in *behaviour* but not necessarily *attitude* (Aronson et al. 2010) whereas persuasion might attempt to influence beliefs, attitudes, intentions, motivations and/or behaviours (Gass & Seiter 2015). This focus on changing attitudes and motivations, rather than simply gaining compliance with an instruction, is also much more in line with approaches described by the therapists.

There are numerous theories of persuasion that attempt to explain if, how and why a change in beliefs, attitudes, behaviour etc. might be achieved. These include e.g. the Theory of Planned behaviour (Armitage & Conner 2001), Conditioning (Rescorla 1971), the Elaboration Likelihood Model (Cacioppo & Petty 1984) and Social Judgement Theory (Sherif & Hovland 1961). The Elaboration Likelihood Model (ELM) was identified as a model that might be suited for application to socially assistive scenarios, and specifically for informing SAR design, based on the associated persuasive strategies (see S1-S6 below) and ‘routes’ to persuasion having significant overlap with behaviours and approaches described by the therapists.

The ELM identifies two routes by which someone receiving a persuasive message (the receiver) from a persuasive source may be persuaded. These are the central route, based on rationale and logic, and the peripheral route, based on stimulus *cues* including a number of *social cues* concerning the source (Petty & Cacioppo 1984). According to the ELM, the processing route taken by a receiver is based on their *elaboration level*; i.e. their motivation and ability to elaborate on the persuasive message. Results from the study with therapists presented in Chapter 2 suggest that the majority of service users would be considered *low elaboration* with regards to the need to do their exercises. As such, it is the low elaboration route of persuasion that is considered in the studies presented here. Of course, there will be some users who would be considered *high elaboration*, and, over time, therapists might hope to change users' attitudes such that they move from being low elaboration to high elaboration with respect to the need to do their exercises. The fact that the model identifies different persuasion strategies to be used in either case improves its potential for informing adaptive/tailored persuasive strategies (and hence SAR design) for maximum effectiveness.

Concerning low elaboration persuasion, the two key source peripheral cues identified by the ELM are credibility and likeability (discussed further below). Similarity between the receiver and the source is also highlighted as a relevant cue, although it is not clear whether this affects credibility and/or likeability specifically (Simons et al. 1970, Wilson & Sherrell 1993, Pornpitakpan 2004).

Defining Credibility

O'Keefe defines credibility as "judgements made by a perceiver (e.g. a message recipient) concerning the believability of a communicator" (O'Keefe 2002). Notably this definition recognises that credibility is subjectively held by the receiver rather than being an objective property of the source. Whilst different sources possess different attributes and abilities, the values assigned to these resides in the receiver, not the source, such that credibility is a perceptual phenomenon.

In an early consideration of what makes a source credible, Aristotle argued the source traits that "induce us to believe a thing apart from any proof of it ...[are] good sense, good moral character and goodwill" (Aristotle 1954). Modern day factor analysis supports his statement; it is now generally accepted that there are three primary dimensions of credibility relevant to the evaluation of a source: *expertise*, *trustworthiness* and *goodwill* (Gass & Seiter 2015). Gass & Seiter (2015) further note that credibility (i) is a multidimensional construct representing a combination of qualities a source is believed to possess, (ii) is a situational/contextual phenomenon affected by audience and setting and (iii) is dynamic and can change over time, even during the course of a single interaction/message. *Extroversion*, *composure* and *sociability* have also been identified as secondary, situation specific dimensions. Scale items used to measure the main expertise, trustworthiness and goodwill dimensions, plus the secondary sociability dimension (identified as being the most relevant to HRI design and informing measures used in the presented studies) are given in Table 3.1.

Expertise	Trustworthiness
Experienced / Inexperienced	Honest / Dishonest
Informed / Uninformed	Trustworthy / Untrustworthy
Trained / Untrained	Open-minded / Close-minded
Qualified / Unqualified	Just / Unjust
Skilled / Unskilled	Fair / Unfair
Intelligent / Unintelligent	Unselfish / Selfish
Competent / Incompetent	Moral / Immoral
Bright / Stupid	Ethical / Unethical
Goodwill	Genuine / Phony Sociability
Cares about me / Doesn't care about me	Good-natured / Irritable
Sensitive / Insensitive	Cheerful / Gloomy
Not self-centred / Self-centred	Friendly / Unfriendly
Concerned with me /	
Not concerned with me	
Has my interests at heart /	
Doesn't have my interests at heart	
Understanding / Not understanding	

Table 3.1: Bipolar adjectives for measuring credibility with a semantic difference scale (Gass & Seiter 2015) implemented as 5-point questionnaire items for the studies presented in this chapter.

Documented strategies for enhancing credibility (Gass & Seiter 2015) that might be applicable to social HRI design include:

- (S1) citing expertise or those of the information source
- (S2) displaying goodwill towards the receiver, i.e. caring about/taking an interest in them
- (S3) improving likability by showing ‘emotional intelligence’ - conveying warmth and immediacy, smiling, remembering people’s names, being polite etc.
- (S4) emphasising similarity between the receiver and the source
- (S5) having the source be introduced by a credible third person
- (S6) adopting a language and delivery style appropriate to the recipient
- (S7) attempting to build trust by demonstrating honesty and sincerity
- (S8) using an assertive style of communication
- (S9) being introduced/endorsed by another source who is already perceived as highly credible

These strategies were also evidenced in the results of the study with therapists presented in Chapter 2, e.g. in SL1's description of how he works differently with two different clients, using different language/delivery style for each and engages in the kind of social interactions appropriate for each of them (needing to have 'fun and a joke' with Patient A but taking a more formal approach with Patient P). Further, therapists referred to such strategies both in providing external motivation through social support/accountability, which would align to the concept of low elaboration persuasion, but also in making themselves credible such that clients engaged with their reasoning. For example, this included rationalising why certain exercises were important in an attempt to improve their intrinsic motivation, which is more aligned with high elaboration persuasion.

3.1.2 Related Work

Reference	Measure of Robot Persuasiveness	Robot Manipulations	Q	P	Q on P
Saunderson & Nejat (2019)*	Compliance with suggestions	Dialogue, gestures, movement	-	-	✓
Lee & Liang (2019)*	Compliance with task request	Performance at a task Request strategy	-	✓	o
Ghazali et al. (2019)*	Compliance with suggestions	Mimicry of movement Affective feedback	✓	o	-
Cruz-Maya & Tapus (2018)*	Concessions in a negotiation	Speed of speech Choice of gestures	o	m	-
You & Robert Jr. (2018)	Trust of & intent to work with	Similarity to participant	✓	-	-
Rossi & D'Alterio (2017)*	Compliance with suggestions	Gaze	-	✓	-
Lohani et al. (2016)	Compliance with suggestions	Social dialogue	✓	✓	✓
Kahn et al. (2015)	Compliance with request to keep a secret	Sociability	✓	✓	✓
Salem et al. (2015)	Compliance with unconventional tasks	Robot errors	✓	o	-
Ham et al. (2015)*	Agreement with a persuasive story	Gaze, gestures	o	✓	-
Ham & Midden (2014)*	Minimising energy consumption	Affective feedback	-	✓	-
Chidambaram et al. (2012)*	Compliance with suggestions	Gaze, gesturing, proxemics	o	✓	✓
Nakagawa et al. (2011)	Time spent/actions on a monotonous task	Touch	✓	✓	o
Siegel et al. (2009)*	Compliance with a request to make a donation	Gender	✓	✓	-
Gockley & Mataric (2006)	Time spent/actions on an exercise task	Engagement in user activity	o	o	✓

Table 3.2: An overview of related HRI studies examining the impact of different robot behaviours in the context of SARs and/or robot persuasiveness. The first column provides a reference and identifies the measure (or proxy measure) of robot persuasiveness employed in the study described. The second column identifies exactly what robot behaviour(s) were being manipulated. The next two columns identify whether those behaviours were found to significantly impact on (Q) any questionnaire measures (i.e. on *perception* of the robot) and (P) *persuasiveness* of the robot as per the identified measure. Finally, the 'Q on P' column identifies whether there was any correlation between participants' perception of the robot (questionnaire responses) and persuasiveness of the robot. Key: ✓ = significant, o = not significant, m = mixed, - = not applicable/not implemented. Note that only those marked * made explicit reference to the concept of persuasion.

Table 3.2 gives an overview of related work, identifying HRI studies which have investigated the impact of different robot behaviours on (i) (objective) user behaviour and (ii) (subjective) user perception of that robot in relevant contexts. Many of these do not refer to the concept of robot persuasiveness directly; including studies which seemingly manipulate behavioural

cues identified by the ELM (e.g. similarity (You & Robert Jr. 2018) and goodwill towards the user (Gockley & Mataric 2006, Kahn et al. 2015)). Of those which do refer to persuasion literature, none refer to the ELM specifically for informing persuasive robot strategies nor understanding persuasion in HRI. Only the most recent works (Saunderson & Nejat 2019, Cruz-Mayo & Tapus 2018) start to describe a potential link between socially assistive robotics and persuasion.

Particularly relevant to this thesis is Gockley and Mataric's early work on using a hands-off mobile robot to encourage physical therapy (Gockley & Mataric 2006). The authors explored whether the behaviour of a very simple, mobile robot could influence participant engagement with exercise tasks typically employed in stroke rehabilitation. They attempted to manipulate perceived robot engagement in the participants' behaviour, with the hypothesis that increased robot engagement would increase the amount of exercise participants would do. The authors ran a pilot study in which participants undertook three open-ended exercise tasks (participants were told to '*repeat this process until you feel that you have exercised your arm enough at this time.*'). Whilst participants were exercising, the 'engaged' robot moved forward and backward in response to participants' exercise movement. For the 'unengaged' robot, movement was completely decoupled from participant behaviour. This specific manipulation failed, however, participants who rated the robot as being more engaged in their activity did do more exercise, providing evidence for the idea that presence of a robot which appears to be interested in the user can influence their behaviour. This is a particularly interesting result given the low social fidelity of the robot platform employed (a Pioneer 2-DX mobile robot). Whilst the paper made no reference to persuasion, it could be argued that 'having an interest in the user' represents the ELM cue of goodwill (Gass & Seiter 2015).

Also employing an open-ended task, Nakagawa et al. (2011) investigated the effect of robot touch on user motivation, demonstrating that active robot touch increased users' number of working actions and working time on the task. Interestingly however, the authors found no correlation between these measures and participants' perception of the robot (feelings of friendliness, authority and trust) as measured by questionnaire. The authors ran a between-subject laboratory study in which participants first interacted with the robot, which either actively stroked their hand, passively touched their hand or did not touch the participant. The robot then asked participants to do a task in a relatively personal/social manner: *'I'd like you to do the following task as well as you can'*. The task was designed to be monotonous and repetitive, and participants were able to end the task at any point.

You & Robert Jr. (2018) investigated the effect of robot-user similarity on trust and intention to work cooperatively with the robot. The authors ran an online, between-subject study in which participants were faced with a hypothetical scenario whereby they would be working collaboratively with a robot on physical tasks in a warehouse. Participants answered a number of questions about work style, and after each one the robot responded that it also chose their answer. The results demonstrated that deep-level similarity (suggested via this simple robot-reported

similarity of answer) increased trust in the robot, and that trust increased intention to work with the robot.

3.1.3 Ethical Considerations

The concepts of robot ethics and ethical robotics are receiving increasing attention from academia, political institutions and the general public. This is evidenced by increasing publications on the topic (see e.g. Vandemeulebroucke et al. (2018) for a review of ethics literature regarding the use of robots in care). There is also a significant number of guidelines emerging, designed to inform robot design and development; a recent review identified 25 distinct sets of such principles for robotics and AI¹.

Considering the philosophical, ethical arguments for the use and design of SARs in e.g. healthcare is not a primary aim of this thesis. However, as part of the responsible approach to innovation taken throughout the work, it is important that ethical implications concerning the robot behaviours designed, tested and discussed in this chapter are considered just as much as any other design, measure of success and/or usability considerations. To this end, the British Standard BS 8611 (BSI 2016) was identified as a practical tool for assessing the potential for ethical harm. The official overview of the standard, provided by the British Standards Institute, is given in the textbox below.

BS 8611 gives guidelines for the identification of potential ethical harm arising from the growing number of robots and autonomous systems being used in everyday life. The standard also provides additional guidelines to eliminate or reduce the risks associated with these ethical hazards to an acceptable level. The standard covers safe design, protective measures and information for the design and application of robots.

Who is this standard for?

- *Robot and robotics device designers and managers*
- *The general public*

Why should you use this standard?

BS 8611 was written by scientists, academics, ethicists, philosophers and users to provide guidance on specifically ethical hazards associated with robots and robotic systems and how to put protective measures in place. It recognizes that these potential ethical hazards have a broader implication than physical hazards, so it is important that different ethical harms and remedial considerations are considered. The new standard builds on existing safety requirements for different types of robots, covering industrial, personal care and medical.

¹<http://alanwinfield.blogspot.com/2019/04/an-updated-round-up-of-ethical.html>

Reviewing the standard instantly identifies the potential for the behaviours discussed in this chapter, both those proposed by the author and those reviewed from existing HRI literature, to be considered as *ethically hazardous*. Specifically, the standard identifies the hazards of *deception* and *anthropomorphization* identifying mitigation strategies as follows.

On deception: *Avoid deception due to the behaviour and/or appearance of the robot and ensure transparency of its robotic nature.*

On anthropomorphism: *Avoid unnecessary anthropomorphization. Clarification of intent to simulate human or not, or intended or expected behaviour. Use anthropomorphization only for well-defined, limited and socially-accepted purposes.*

In both cases, *user validation* and *expert guidance* are given as tools for verification/validation for assessing and mitigating such risks.

The work presented in this chapter already encapsulates some expert guidance, as it is informed by the study with therapists presented in Chapter 2. In addition, all of the studies presented in this chapter are designed to include an element of user validation. Specifically, experimental measures concerning perceived deception in and acceptance of, the demonstrated robot behaviours were included. Further, Studies 2 and 3 of this chapter are specifically concerned with investigating the potential impact that better complying with these mitigation strategies might have on the effectiveness of socially persuasive robot behaviours. To the author's knowledge, this initial, practical consideration of the potential 'cost' of designing more ethical robot behaviours, specifically informed by a published standard, is novel in the field of social HRI.

3.1.4 Research Questions and Overview of Studies Presented

As with Chapter 2, the studies presented in this chapter are grounded in the SAR application of supporting therapy engagement, particularly limiting the consideration of acceptability noted in RQ4, to this application domain. However, the aim for the other research questions is to inform SAR and social HRI more generally, such that they can be summarised as follows:

RQ1 Can HHI credibility-boosting/persuasive strategies increase objective robot persuasiveness?

RQ2 Can HHI credibility-boosting/persuasive strategies increase subjective/perceived robot credibility and/or likeability?

RQ3 Is there a correlation between questionnaire measures designed to measure *perception* of a robot and objective measures (of user behaviour) designed to measure robot *persuasiveness*?

RQ4 Are socially persuasive robot behaviours acceptable in the context of assistive HRI? Are they considered e.g. *genuine*, or *deceptive*?

RQ5 Would designing socially persuasive robot behaviours that in a way that minimises ethical risk have any impact on their potential effectiveness?

CHAPTER 3. PERSUASION AS A MODEL FOR SOCIALLY ASSISTIVE HRI

Study	Medium	Manipulations	Perception and Subjective Measure(s)	Persuasion Measure	RQ's
1	Laboratory	Socially persuasive behaviour strategy Between-subject: - Goodwill - Similarity - Expertise - Control (neutral social interaction)	Credibility (Gass & Seiter 2015) Likeability (Bartneck et al. 2009) Responsibility ascription Relationship development (Hall et al. 2014) Genuiness (goodwill, similarity only) Acceptability/Deception	Exercise repetitions	1, 2, 3, 4
2	Online	Source of expertise presented by the robot Within-subject: - Robot's own expertise - Therapist's expertise - Control (no expertise)	Credibility (Gass & Seiter 2015) Likeability (Bartneck et al. 2009) Responsibility ascription Relationship development (Hall et al. 2014) Robot preferences		2, 5
3	Online	Design of social dialogue Within-subject: - Anthropomorphic (less ethical) - Ethical (less anthropomorphic) - Control (no social dialogue) Between-subject: - Priming of potential for deception	Credibility (Gass & Seiter 2015) Likeability (Bartneck et al. 2009) Responsibility ascription Relationship development (Hall et al. 2014) Robot preferences Acceptability/Deception		2, 4, 5

Table 3.3: Overview of experimental studies on persuasive social robot behaviour presented in this chapter.

Three experimental studies were used to address these research questions. An overview of these studies describing the manipulations and measures employed, as well as which research questions are addressed in each study, is given in Table 3.3. As shown in Table 3.3, a number of the measures were used across all of the studies. These are described below and presented in full in Appendix B. The measures on responsibility ascription and relationship development were administered using 5-point Likert response scales, and were presented after the main credibility and likeability measures. Study specific measures are detailed within the respective subsections below.

Credibility

Robot credibility was measured using questionnaire items designed to measure credibility of a human source; with 5-point Likert question items arranged in subscales of expertise, trustworthiness, goodwill and sociability (as presented in Gass & Seiter (2015), adapted from McCroskey & Teven (1999) and McCroskey & Young (1981)). The question item descriptors are given in Table 3.1.

Likeability

Robot likeability was measured using the likeability scale of the Godspeed questionnaire (Bartneck et al. 2009). Other items from the Godspeed questionnaire were not included due to significant overlap with the credibility measure.

Relationship Development

Relationship development questions were worded as below, taken from a previous study

investigating engagement in HRI (Hall et al. 2014), and were included based on the importance of the therapist-patient relationship in therapeutic exercise engagement discussed in Chapter 2:

To what extent do you feel [you / the patient] developed a relationship with the robot?

To what extent do you feel the robot developed a relationship with [you / the patient]?

Ascription of Responsibility

Ascription of responsibility to the robot offers an applied/tangible measure of credibility, however, this is limited given participants do not *actually* have to work with the robot as part of a therapy programme. This was worded as follows:

For the laboratory study:

Imagine you were undergoing a therapy regime where you had to do exercises every day, and you had this robot at home to help you in-between visits from your therapist. How much responsibility do you think the robot should hold for helping with your exercise regime?

For the online studies:

How much responsibility do you think this robot would hold for monitoring Katie's engagement with her home exercises?

Preferences

For the online, within-subject online studies, participants were asked to identify which of the robots they found most motivating and would rather work with:

Which robot do you think was most motivating, and why?

Which robot would you rather work with, and why?

3.2 Study 1: ELM for Socially Assistive Robotics

The aim of this study was to investigate whether persuasive strategies employed in human-human interaction (HHI):

- (i) can increase the persuasiveness of a social robot, measured objectively through participant behaviour
- (ii) can increase credibility/likeability of the robot, measured by questionnaire
- (iii) are perceived as deceptive and/or acceptable to participants

The experimental protocol, whilst laboratory based, was grounded in the context of therapeutic exercise instruction and encouragement, as considered and therefore informed by the work in Chapter 2. Few previous works consider this link between persuasiveness and assistance, and none have specifically investigated the applicability of the ELM.

3.2.1 Methodology

A four condition, between-subject, wizard-of-oz (WoZ cf. (Riek 2012)) laboratory study was designed using the social robot Pepper². Three of the conditions were designed to reflect persuasive strategies derived from the ELM: goodwill, towards the participant, similarity between the robot and the participant and expertise related to the task scenario. The fourth condition was designed to represent a control of neutral social interaction. Full details regarding the conditions are given in Section 3.2.3.

An exercise session interaction scenario was designed in order to give the study real world context and applicability, whilst representing a low elaboration scenario for participants. Specifically, the robot asked participants to do repetitions of a wrist turn, a simple exercise designed to treat Tennis Elbow³. To minimise relevance/inherent motivation arising from the exercise itself, participation criteria required '*no mobility issues affecting wrist movement in either arm*'. The study was advertised simply as a study on social robots for exercise, concerned with '*how such a robot might behave and how different robot behaviours are perceived/which ones are preferred by users*'. The study was approved by the Faculty of Science ethics committee of the University of Bristol.

A total of 92 participants were recruited through online/poster advertisements and on-campus leafleting on a rolling basis across two weeks of experimental sessions. Participants included 41 males, 50 females and 1 participant of undisclosed gender with a categorical age distribution as shown in Figure 3.1. Data for 2 participants were disregarded due to technical errors. Participants were allocated to conditions as follows: control (N = 22, 12 female), goodwill (N = 28, 15 female, 1 undisclosed), similarity (N = 20, 11 female) and expertise (N = 20, 10 female). Allocation to condition was random except for 7 male participants toward the end of the recruitment phase who were assigned to the goodwill condition to account for a gender imbalance in that condition (resulting from the rolling recruitment/randomisation process). Participants were offered a £5 Amazon voucher for taking part in the experiment.

3.2.2 Experimental Measures

Robot persuasiveness was measured objectively by the number of wrist turn repetitions completed by participants. Exercise duration (time spent voluntarily exercising with the robot following the pre-exercise dialogue) was also recorded, and participant exercise speed was approximated post-hoc by dividing number of repetitions by this exercise duration. Perception of the robot/subjective measures included credibility, likeability, relationship development, acceptability and deceptiveness as discussed previously. The credibility and likeability measures were administered before *and* after the exercise session interaction in order to record participants baseline perception/expectations of the robot.

²<https://www.softbankrobotics.com/emea/en/robots/pepper>

³<https://www.versusarthritis.org/about-arthritis/conditions/elbow-pain/>

3.2. STUDY 1: ELM FOR SOCIALLY ASSISTIVE ROBOTICS

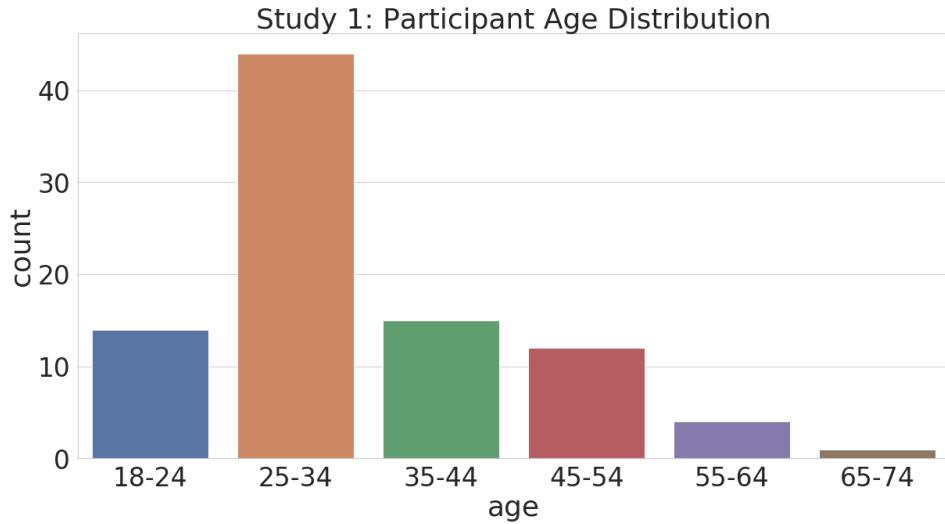


Figure 3.1: Age distribution for Study 1 participants.

Specific to this study, a question on genuineness was included because HHI literature suggests a lack of genuineness may reduce persuasiveness. Specifically, if the source is perceived to be simply feigning interest in order to be persuasive then such strategies will not be effective (McCroskey & Teven 1999). This question was only included for the goodwill and similarity conditions, as it was focused on the genuineness of those behaviours rather than of the robot overall. The question read as follows, and was administered using a 5-point Likert response scale, inline with the other measures:

“The robot you saw attempted to show some [goodwill / similarity] towards you by [asking how you felt about being here and doing the exercises / showing an interest in your responses and reacting accordingly by suggesting it had the same exercise preferences that you do]. How genuine did you perceive that behaviour to be?”

Participants were also asked whether they perceived the robot to be deceptive. A brief explanation of why this might be the case was provided, to account for any potential lack of understanding regarding social robots and their capabilities:

“There is a growing concern amongst some roboticists that social robot behaviours are deceptive as robots do not and cannot feel emotions, nor do they have any real interest in the person they are interacting with. Do you think the robot you saw today was deceptive? If so, do you feel that deception was acceptable?” (Yes - deceptive but acceptable / Yes - deceptive and not acceptable / Not deceptive / Not sure)

After completing the exercise session and post-hoc questionnaire, participants were invited to take part in a brief interview utilising the following topic guide:

1. Describe the robot exercise instructor; any particular likes/dislikes

Designed to identify any high-level differences in perception of the robot across condi-

tions, and whether the manipulated behaviours featured in participants descriptions of the robot/opinion of the interaction.

2. Reasoning behind answers to question on genuineness (similarity and goodwill conditions only)
3. Reasoning behind answers to question on deception
4. Revisit of above/additional comments after debrief

The debrief explained the study aim of investigating persuasiveness and the behaviour manipulations employed; the researcher probed whether this might change participants' answer to the genuineness and/or deception questions, or whether they had any additional comments to make.

3.2.3 Experimental Conditions

The experimental conditions were designed to demonstrate robot-participant similarity, robot goodwill towards the participant and task-relevant robot expertise through robot-initiated dialogue. All conditions were designed around the same dialogue/interaction pattern; in each case the robot asked the participant three questions requiring a response, before introducing the exercise task. All dialogue concerning the exercise task (instructions, encouragement etc.) and overall dialogue duration was identical across conditions. The dialogue employed for each condition is shown in Table 3.4.

In the similarity condition the robot suggested to the participant that they should compare preferences for scheduling exercises, and asked them three questions selected and adapted from the Stroke Exercise Preference Inventory (Bonner et al. 2016). Whichever answer the participant selected, the robot indicated it had also chosen that answer, based on the procedure employed by You & Robert Jr. (2018). In the expertise condition, the robot introduced itself as being programmed by physiotherapists, asked questions concerning the participants previous experience with therapy and provided a number of facts about the exercise to be done/the condition it was designed to treat. This information was taken from public NHS⁴ and Arthritis Research UK⁵ self-help material. In the goodwill condition, the robot asked questions designed to demonstrate an interest in the participants' feelings toward the session, and responded with an emotionally-matched response. Finally, the control condition was designed to be as neutral as possible, with the robot providing some factual information about a number of topics unrelated to the interaction scenario and asking the participant some questions pertaining to those. Beyond this initial dialogue, each interaction followed a set procedure as described below:

⁴<https://www.nhs.uk/conditions/tennis-elbow/>

⁵<https://www.versusarthritis.org/about-arthritis/conditions/elbow-pain/>

Control Condition Dialogue
I was designed and built by Softbank Robotics in Paris. I am 1.2 metres tall and weigh 28 kilograms. Have you worked with a robot like me before?
Ok. Car travel is the most common mode of transport in Bristol. However, Bristol is also one of the most prominent cycling cities in the country. How did you get here today?
I see. This summer was one of the hottest on record in the UK. Sometimes Bristol was hotter than Paris. What is the weather like today?
Similarity Condition Dialogue
Before we get started, let's compare our preferences for scheduling exercises. Here are some questions about exercise. Please tell me your opinion and we can compare it to my answers. First, if you had to choose one or the other, is it better to exercise alone or with others?
I also chose [participant answer].
Next, if you had to choose, is it better to exercise whilst watching tv or listening to music, or is it better to concentrate only on the exercise?
I also chose [participant answer].
Finally, if you had to pick one or the other, is it better to exercise outdoors or indoors?
I also chose [participant answer]. It seems like we have similar ideas about exercising.
Expertise Condition Dialogue
I have been programmed by physiotherapists who specialise in exercise for pain relief. Have you ever worked with a physiotherapist?
(Y) Was that very recently?
(N) Have you ever worked with a personal trainer?
Ok. Today we are going to do an exercise designed to treat Tennis Elbow. Tennis Elbow is caused by a strain to tendons in your forearm. It can be easily treated and should ease within two weeks. Have you ever suffered from tennis elbow?
I see. Tennis elbow is a common musco-skeletal condition. It's estimated that as many as one in three people have tennis elbow at any given time. It usually affects adults and is more common in people who are 40 to 60 years of age.
Goodwill Condition Dialogue
I'm pleased to meet you and looking forward to working together. Before we start I would like to get to know you better, so I'm going to ask you some questions. How do you feel about being here today?
(P) Great, I'm glad to hear that! I'm sure you will enjoy the session.
(Neg) I'm sorry to hear that, hopefully you will enjoy the session.
(Neut) I understand. Well I hope you will enjoy the session.
And how do you feel about working with a robot?
(P) That's good to hear, we'll definitely have fun together today then.
(N) I can understand that, but I hope we can still have fun together today.
As you know, today we are going to do some exercise, do you enjoy exercising?
(P) That makes sense, this session will be easy for you then.
(N) That's understandable, this exercise is quite easy though so hopefully won't be too bad.

Table 3.4: Pre-exercise robot dialogue employed in each experimental condition, designed to manipulate perceived robot similarity, expertise and goodwill compared to the control condition.

3.2.4 Experimental Procedure

Participants were first given an information sheet to read and asked to complete an initial consent form before providing demographic information. Regarding the open-ended exercise task, the information sheet specifically stated:

“Pepper will interact with you and guide you through an open-ended wrist turning exercise. If/when you stop exercising, the robot will note that you’ve stopped and ask if you want to finish.”

Participants were then led into the experimental area and introduced to Pepper, which was turned on but in ‘sleep’ mode. As shown in Figure 3.2, the experimental area was designed to shield the participant from external observers. This was designed to minimise any observation/demand effects which might influence their behaviour, as well as to mask the WoZ nature of the study. The researcher then explained that the robot was in an un-responsive sleep mode, and that before starting the exercise session the participant could take some time to familiarise themselves with the robot (e.g. by visually inspecting it, touching it, moving the head and hands). This was encouraged in order to give participants at least some baseline experience and time to reflect on and inform their expectations of the robot (then captured via their pre-hoc questionnaire responses). Participants were asked to complete the pre-hoc questionnaire when ready, then to stand in a pre-defined spot marked on the floor and verbally indicate to the researcher that they were ready to start the exercise session. Regarding the exercise task, the researcher again explained that the exercise was open-ended, using the same phrasing as the information sheet. The researcher then left the experimental area and waited for the participant to indicate they were ready to begin.

On launching the experimental script, Pepper displayed its standard start-up animation sequence. The wizard then followed a set protocol for the exercise session interaction as shown in Figure 3.3. The protocol accounted for participants ceasing to exercise, doing the exercise incorrectly/doing some other unexpected behaviour and asking the robot additional questions during the task. The wizard also manually logged each repetition done by the participant, with encouragement then being automatically given at 1, 3, 5, 7, 10 and 15 repetitions. At 20 and 25 repetitions the robot moved its head with no speech, to suggest it was still active. Encouragement was given more frequently at the beginning of the exercise to ensure participants were confident that their technique was correct and that the robot was watching/reacting to their actions.

The WoZ control setup included pre-programmed options to have the robot indicate that the participant’s exercise didn’t look correct (for responding to unusual participant behaviour or purposeful error) and to state it didn’t know how to respond (in the case of unexpected/additional participant questions). If/when the participant stopped exercising, the robot asked whether the participant wanted to finish. If they said yes then the completion message was played, otherwise another encouragement message was played and the wizard continued to log repetitions. The exercise task was capped at 30 repetitions (to minimise the potential for any repetitive strain injury/discomfort), upon reaching which the robot automatically said the participant had done

3.2. STUDY 1: ELM FOR SOCIALLY ASSISTIVE ROBOTICS

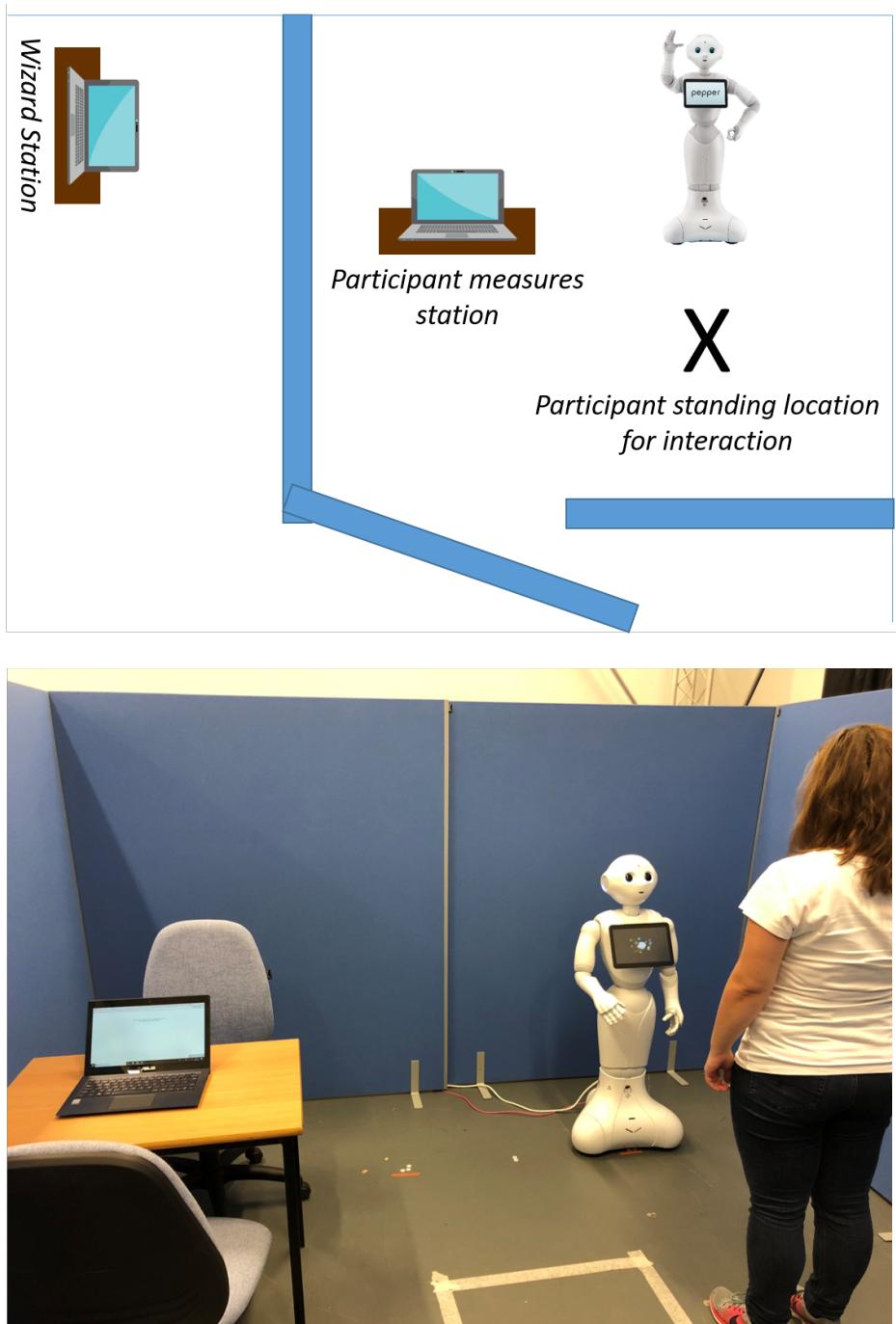


Figure 3.2: Diagram of the experimental room layout and photograph showing the enclosed interaction space. **Note that the laptop shown in the photograph was used to collect participant questionnaire responses only;** the wizard station was external to the interaction space and at some distance from the enclosed area.

enough, followed by the completion message. In all cases the robot then returned to sleep mode, displaying its standard shutdown animation sequence.

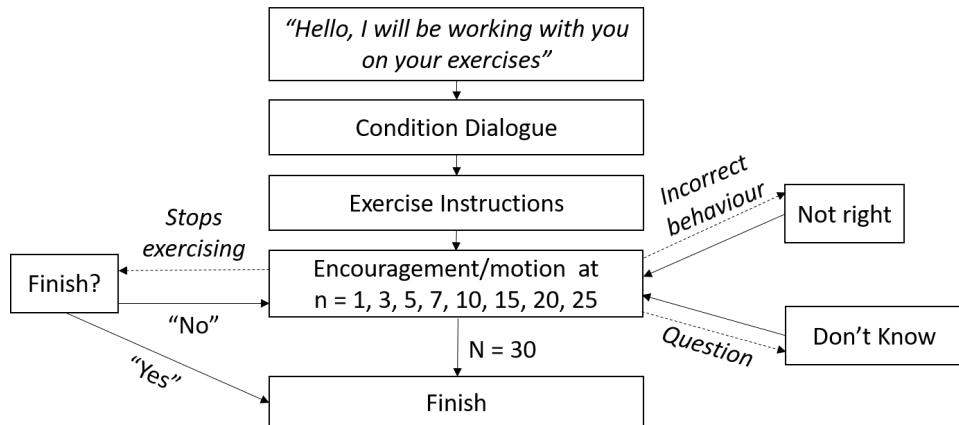


Figure 3.3: Stages of the exercise session interaction highlighting wizard protocol for generating dialogue and responding to participant behaviour.

3.2.5 Results

Exercise Behaviour (Robot Persuasiveness)

Exercise speed (calculated as the number of repetitions completed divided by time spent exercising with the robot) was not uniform across participants, varying from 0.05 repetitions/second to 0.41 repetitions/second ($M = 0.17$, $SD = 0.09$). ANOVA analysis suggests that this variation in exercise speed was not significant between groups ($F(3,88) = 0.321$, $p = 0.810$); suggesting it was unaffected by the experimental manipulation. As such, exercise speed could be considered as a covariate when analysing number of repetitions (to account for the potential that participants who do repetitions very quickly are more likely to do lots of repetitions, and vice versa). ANCOVA analysis was therefore used for analysis of the repetition data. The data are not normally distributed but exhibit homogeneity of variance as determined by Levene's test, thus making ANCOVA/ANOVA appropriate for the following analyses.

There was a statistically significant difference in the number of repetitions completed by participants between groups, as determined by one-way ANCOVA analysis, accounting for exercise speed as a potential confounding variable ($F(3,88) = 8.123$, $p < 0.001$). The effect size of experimental condition was only small (0.225) but larger than that of the confounding exercise speed variable (0.128). Figure 3.4 shows the distribution of participant repetitions for each condition, clearly demonstrating a ceiling effect in the goodwill and similarity conditions with a large number of participants completing the maximum 30 repetitions, suggesting they may have continued further had that limit not been imposed. A Bonferroni post-hoc test revealed that the number of repetitions was significantly higher in the goodwill ($M = 24.9$, $SD = 8.2$; $p < .001$) and similarity ($M = 25.3$, $SD = 7.5$; $p < .001$) conditions compared to the control condition ($M =$

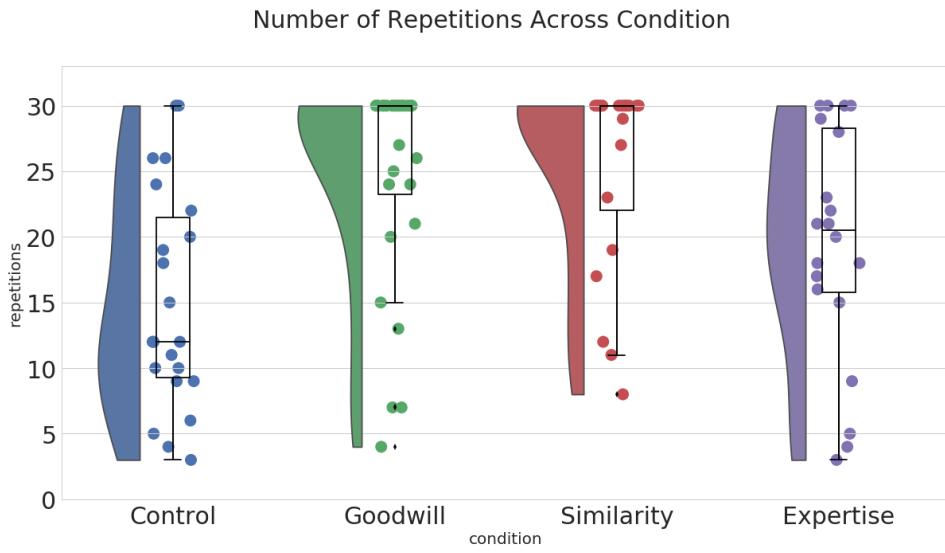


Figure 3.4: Boxplot and distribution of wrist turn repetitions done by participants in each experimental condition. The goodwill and similarity conditions show a significant ceiling effect.

15.1, SD = 8.4). There was no significant difference between the expertise ($M = 19.5$, $SD = 8.9$; $p = .345$) and the control condition. There was also no significant difference between the expertise and goodwill/similarity conditions.

Robot Credibility and Likeability

Post-hoc credibility and likeability were not found to vary significantly between groups. Specifically, one-way ANOVA analysis of questionnaire subscales returned the following results: expertise ($F(3,89) = .786$ $p = .505$), trustworthiness ($F(3,89) = 2.599$, $p = 0.057$), goodwill ($F(3,89) = 2.322$, $p = 0.081$), sociability ($F(3,89) = .831$, $p = .480$) and likeability ($F(3,89) = 1.176$ $p = .324$). A paired samples t-test comparing the within-subject pre- and post hoc questionnaires for all participants across all conditions demonstrated a significant increase in the goodwill ($t = 5.905$, $p < .001$) and sociability ($t = 3.237$, $p = .002$) subscales of the credibility questionnaire. Likeability ($t = 6.089$, $p < .001$) also significantly increased. ANCOVA analyses showed there was no difference in these within-subject shifts between groups.

Neither the likeability measure nor any subscale of the credibility measure, was found to significantly correlate with the number of repetitions participants completed or the time spent exercising. In summary, there was (i) no significant difference in participants' *perception* of the robot's persuasiveness across conditions and (ii) *no correlation* between participants' perception of the robot and the extent to which they appear to have been persuaded by it.

Ascription of Responsibility

Ascription of responsibility to the robot did not vary significantly across groups. In addition, mirroring the credibility and likeability results, responses to this question did not correlate

with the number of repetitions participants did. Overall, after completing the exercise session, participants indicated they would not ascribe much responsibility to the robot ($M = 2.43$, $SD = 1.16$).

Relationship Development

Relationship development to/from the robot did not vary significantly between groups and similarly did not correlate with the number of repetitions participants did. A paired samples t-test demonstrated a significant difference between answers to those two questions ($t = 3.756$, $p < .001$). Specifically, average relationship development to the robot ($M = 3.20$, $SD = 1.06$) was scored higher than relationship development from the robot ($M = 2.82$, $SD = 1.08$). Further, whilst answers to those two questions were correlated, this correlation was only moderately strong ($r = 0.602$). This would suggest participants recognised a clear difference regarding their development of a relationship to the robot, versus the robot's development of a relationship with them.

Genuineness and Acceptability / Deception

Concerning the genuineness of dialogue in the similarity and goodwill conditions, participant answers did not vary significantly between the two groups as measured by a Mann-Whitney test ($Z = -1.20$, $p = 0.23$). Across both conditions, answers to the genuineness question were very mixed, with approximately half of participants assigning a low score (1 or 2 out of 5) but a quarter giving the maximum score of 5/5. Once again, responses to this question did not correlate with the number of repetitions participants did. Concerning deception and acceptability, the distribution of answers was somewhat unexpectedly similar across all conditions, including the control. The majority of participants considered the robot not deceptive or deceptive but acceptable. The distribution of answers to both of these questions are given in Figure 3.5. Participant reasoning behind these answers (collected initially via questionnaire and further discussed in the post hoc interview) are discussed below.

Qualitative Data

All participants chose to take part in the post-exercise interview. Interview data was coded for emergent themes concerning descriptors used to describe the robot, likes/dislikes, acceptability of behaviour and/or deception and, for the goodwill and similarity conditions only, genuineness of behaviour. These themes are presented in Figures 3.6 to 3.11.

Descriptors

Across all conditions, participants suggested they perceived the robot to be friendly and clear, but expressed some uncertainty over the extent to which the robot was really tracking their exercise performance and therefore to what extent it could provide task-specific feedback around correct execution of the prescribed exercise. Potentially offering a form of positive manipulation check, the goodwill and similarity conditions evoked more descriptions of being motivating, the

3.2. STUDY 1: ELM FOR SOCIALLY ASSISTIVE ROBOTICS

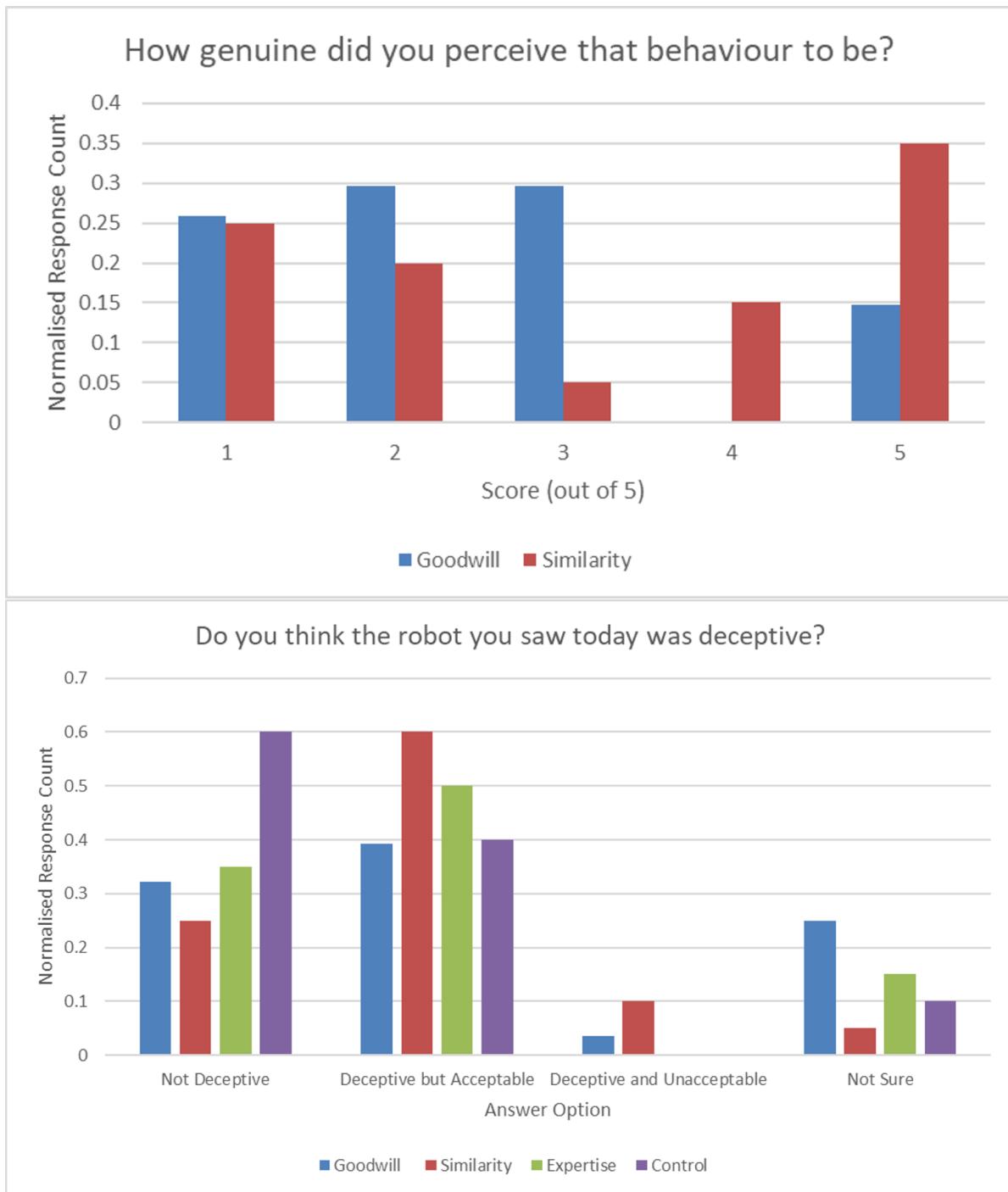


Figure 3.5: Frequency count (normalised against the number of participants in each condition) of participant responses to the questions on (i) genuineness of the robot's behaviour, for the goodwill and similarity conditions only (scored on a 5 point Likert scale) and (ii) whether participants felt the robot was deceptive, across all conditions.

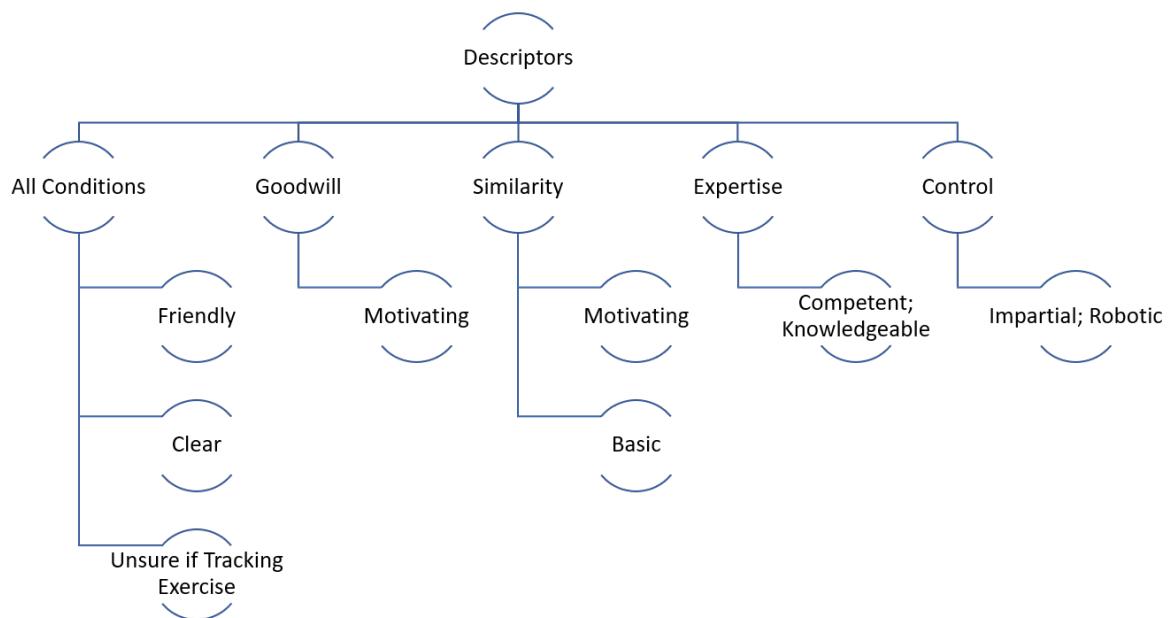


Figure 3.6: Emergent themes concerning how participants described the robot in each condition. Whilst all participants generally described the robot as being friendly, only those in the goodwill and similarity conditions tended to describe the robot as motivating. Similarly, the expertise robot was also uniquely described as competent/knowledgeable; suggesting the desired manipulations of ELM persuasive strategy were successfully implemented for each condition.

expertise condition of being competent or knowledgeable and the control condition of being fairly impartial or robotic.

Likes

Across all conditions participants typically expressed either an overall like (or dislike, see Dislikes below) for the robot's attempts to engage in some social interaction and for elements of the robot's design (e.g. voice, humanoid shape). Again somewhat reflecting the condition manipulations, in the goodwill and similarity conditions, participants typically identified that they liked the encouragements given by the robot. In the expertise condition this did not typically come up, although participants did express a liking for the expertise demonstrated by the robot. Neither feature was described by participants of the control condition. This is particularly interesting given the encouragement dialogue during the exercise task itself was *identical* across all conditions, possibly suggesting that the initial conditioned dialogue of the goodwill and similarity conditions somehow influenced the effectiveness/impact of those encouragements given later during the task.

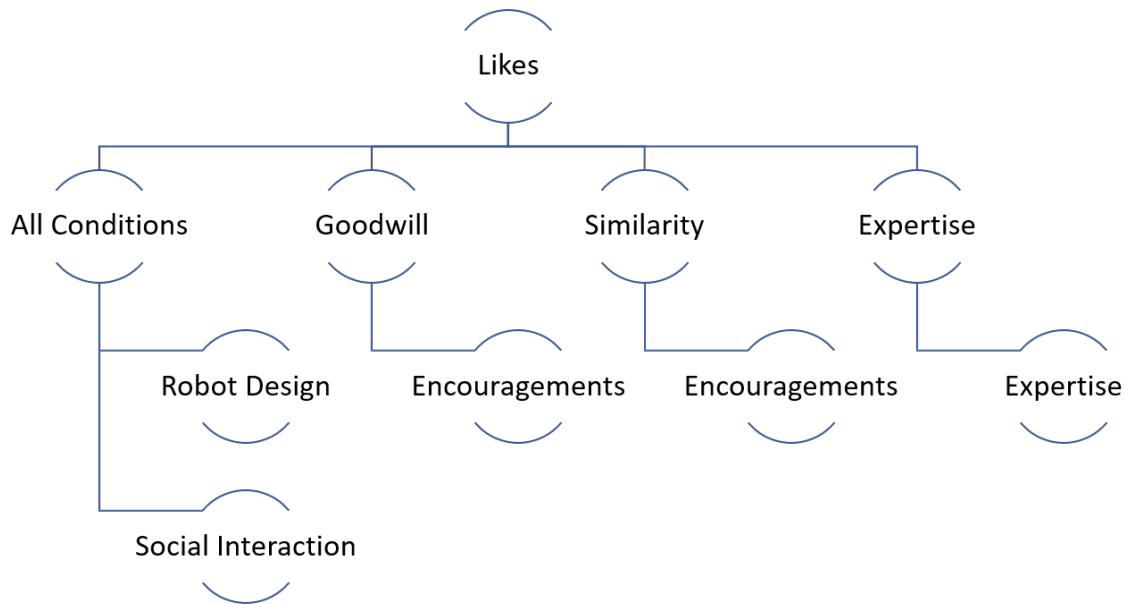


Figure 3.7: Emergent themes concerning what (if anything) participants liked about the robot in each condition. A number of participants in all conditions referred to liking the social interaction the robot engaged in, and/or highlighted some platform specific design features they liked such as the robot's voice; although other participants highlighted these same features as dislikes (see Figure 3.8). Only those in the goodwill and similarity conditions tended to specifically mention liking the encouragements provided by the robot, even though the encouragements given during the exercise itself were *identical* across conditions. Again suggesting positive manipulation of the desired cue, participants in the expertise condition reported liking the expertise demonstrated/information provided by the robot.

Dislikes

There was not one consistent theme that emerged across all of the conditions, although there was a lot of feedback regarding a need for more interaction, either social, task specific or for encouragement. Interestingly these were also seemingly somewhat aligned to the conditions, i.e. in the expertise condition participants generally referred to there not being enough technical feedback, whereas in the goodwill and similarity they referred to there not being enough encouragement and finally in the Control condition they commented on the very limited/surface level social interaction. Certainly for the goodwill, similarity and expertise conditions, particularly when combined with the descriptor results, this might suggest how the initial conditioned dialogue set up an expectation regarding how encouraging (in goodwill and similarity) or technically proficient (in expertise) the robot would then be during the exercise task.

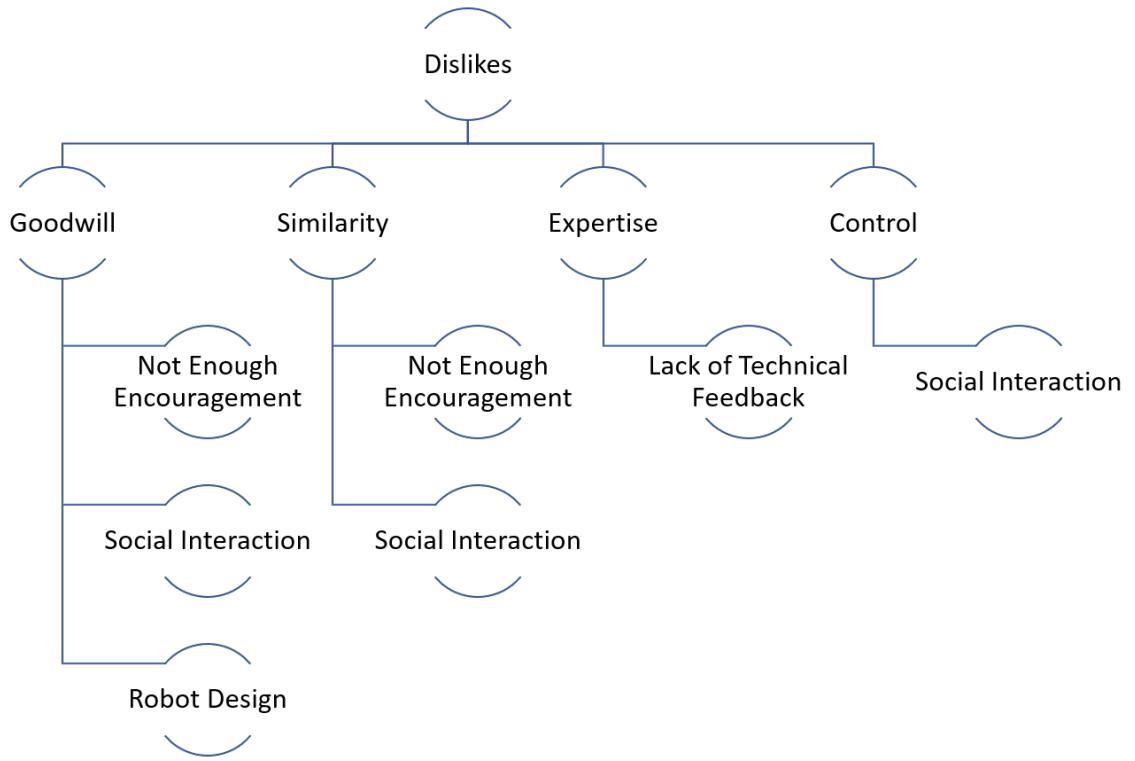


Figure 3.8: Emergent themes concerning what (if anything) participants disliked about the robot in each condition. In all but the expertise condition, a number of participants referred to disliking the social interaction the robot engaged in, even though this was equally highlighted by other participants as being something they liked (see Figure 3.7). Another common theme across conditions was that the robot should interact more, although in the goodwill and similarity conditions this tended to be in terms of encouragement and/or better social interaction whereas in the expertise it was more about the lack of technical feedback regarding performance of the exercise.

Genuineness

Across both the Goodwill and Similarity conditions, participants expressed the idea that the interaction felt fairly genuine, even though they knew the robot did not really *feel* the affective concern/interest expressed:

[G18]: “*I knew he didn’t really care, but when he said it it did kind of feel genuine*”

[G9]: “*I put it on the genuine side, but not really genuine because I know obviously it’s not...but it does feel like it*”

[S17]: “*Obviously I knew it wasn’t true but there was something at the back of my mind saying he was, that was really sort of nice, that he was caring somehow*”

In the similarity condition most participants expressed that at the time of interaction they were fairly certain that (or at least began to question whether) the robot was simply mirroring their answers, but participants were then split on how they felt about this. Specifically, some felt

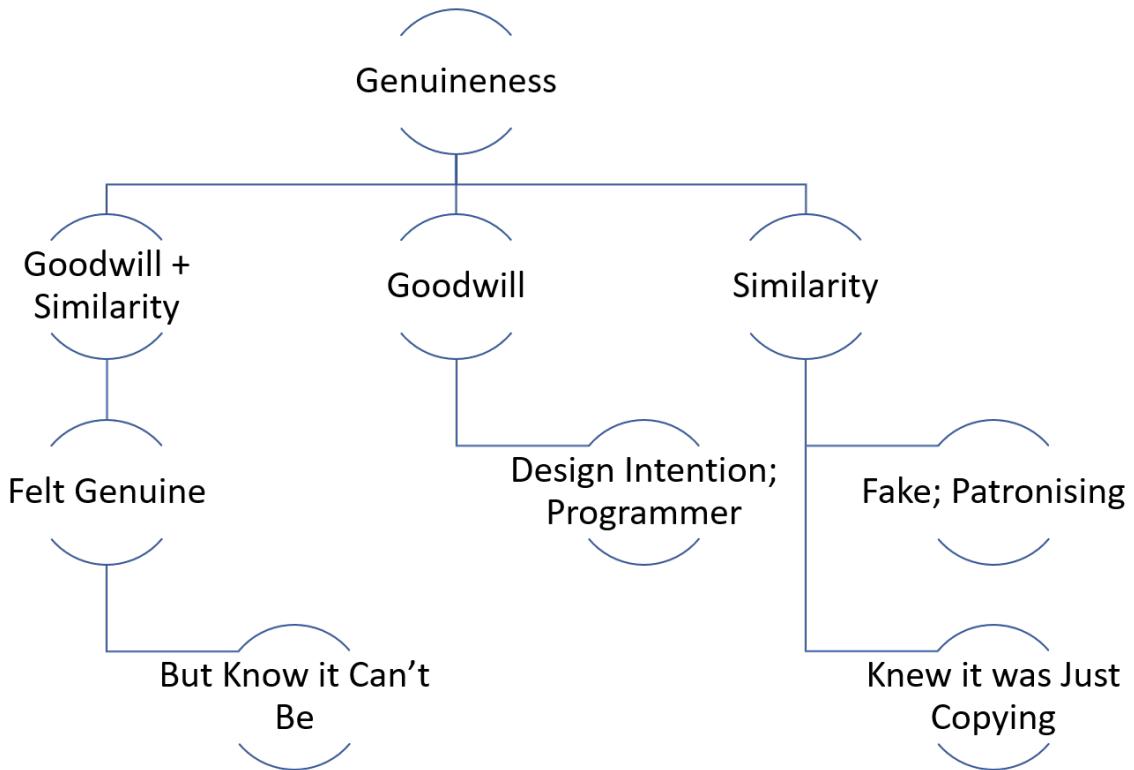


Figure 3.9: Emergent themes concerning discussion on how participants of the goodwill and similarity conditions answered the question '*Did you perceive the robot's behaviour to be genuine?*' Across both conditions a number of participants identified that the robot's interactions, affect or concern *felt* genuine even though they perhaps *knew* that they couldn't be. Most participants in the similarity condition seemed to realise that the robot was simply mirroring them, and for some was somewhat negatively described as fake and/or patronising. In the goodwill condition, participants often also brought up the genuineness of the robot's purpose and the intentions of the programmer/designer etc. as linked to that.

it was completely fake/disingenuous and therefore potentially patronising, something not really seen in the goodwill condition:

[S21]: "*I didn't like it... the response times indicated that she was just going to agree with me whatever, so I didn't get the impression that was her opinion... it felt a little fake... I think it probably would have been better had she kind of disagreed with me... and given a bit more of a justification as to why she thinks that rather than just instantly agree... that would have made it a little bit more like we were on the same page*"

Participants in the goodwill condition specifically commented on the genuineness of the intention *behind* the affective demonstrations, typically linking to that of the humans who designed/implemented the robot:

[G12]: "*I felt like it was genuine but I'm also very aware that somebody else programmed it to be genuine but I'm ok with that because I feel like whoever had made the program in the first*

place did want the person who was doing the training to feel comfortable and cared about... it is the intention behind it"

[G20]: "*Well I think it comes from a genuine place even if it's not yet probably genuine in the sense of the robot but the fact it's been programmed to do that does come from sort of a genuine place of care*"

Building on this idea of whether the genuineness was really about the social behaviours demonstrated by the robot directly, and how this *felt*, or rather how those behaviours fitted in to the larger picture of what the robot was designed/who had designed it (see further on this under *Acceptability of Behaviour*), participants often expressed some difficulty in understanding or applying the concept of *genuineness* in the context of a social robot:

[G5]: "*I'm not sure how you would define a genuine interaction with a robot, what does genuine mean? In a sense it's not good or bad because it's as good or bad as the people that created this as an experience so I can't put this burden on the robot, it's not the robot's fault*"

Deception

Across all conditions, and particularly relating to the social behaviours of the robot, three key reasons for it being considered not deceptive, or deceptive but acceptable were (i) the robot is simply following its programming:

[G7]: "*It's not deception because it's been programmed to do a certain thing, and so it's not deceiving anyone*"

and building on this (ii) just as it is incapable of *feeling* emotions it is similarly incapable of *purposefully* deceiving you:

[G26]: "*Not having emotions means that it can't choose to be deceptive or not...it's just, it's been programmed to respond in a certain way to certain stimulus, it's just doing what it's told*"

and (iii) participants were never deceived as to the reality of the robot's nature, i.e. always being aware that it was '*just a robot*' when it was displaying social behaviours (even if they somewhat '*bought in to*' those behaviours at the time):

[S21] "*I did say it was deceptive on the form but because I feel like it's a program, a pre-programmed response. At the end of the day I realise it's a computer, well it's a robot and it's pre-programmed so to some degree it's deceptive... but I expect that*"

However, whilst the deception question was very much introduced in regard to the robot's social behaviour, across all conditions participants also commented on the potential for deception with regards to how much the robot was really tracking their exercise performance, given that it was actually giving some feedback suggesting it was doing so:

[S13]: "*I put deceptive and unacceptable... going back to the lack of genuineness and it felt a bit fake... trying to say these things to me that weren't necessary but then the things I thought it would have corrected me on, it hadn't. So it's almost like you start off really friendly but then when I do the exercise wrong you don't tell me it's wrong so that's why I think it felt deceptive because it didn't seem to have my best interests at heart*"

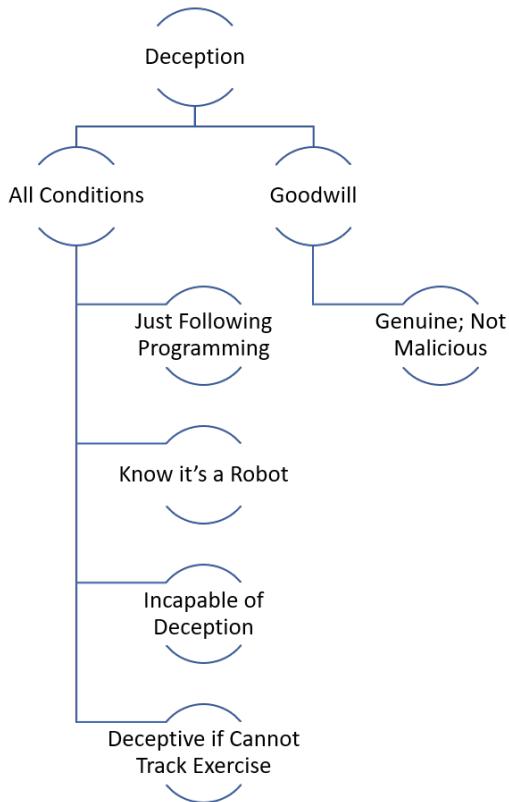


Figure 3.10: Emergent themes concerning discussion on participants answer to the question '*Did you perceive the robot you saw today to be deceptive?*' Across all conditions reasons for the robot *not* being deceptive centered on its robotic nature (either making it obvious it's a robot or rendering it incapable of deception) in addition to the idea that it was just following its programming. An unexpected theme however was the potential for deception regarding the robot suggesting it was *watching or monitoring* the participant's exercise behaviour if in reality it wasn't. In the goodwill condition specifically, participants also referred to a lack of malicious intent/how they felt the interaction to be genuine.

[G27]: “*A real doctor would... maybe you winced or something and then that real person would flag it, but if the robot does not have that kind of sensing capability then... maybe it would attempt to provide the best answer to the best of its capabilities...so I doubt it is intentionally deceptive but the feedback it provides can be*”

Moving towards acceptability of potentially deceptive behaviours, many participants felt somewhat uncomfortable calling the goodwill and similarity behaviours deceptive because of the feeling that they were ‘genuine’ as discussed above) or at least not *malicious* in their intent:

[G15]: [*The last question asked about deception, it said the robot can't really feel excited to meet you and things like that. How did you feel about that?*] “*No not at all... you know what, I felt that it did... I did I really felt that it was talking to me, made me happy, made me understand. It was absolutely brilliant*” (G15)

[G11]: “*With deceptive you kind of feel like it's sort of malicious, and because it seemed very*

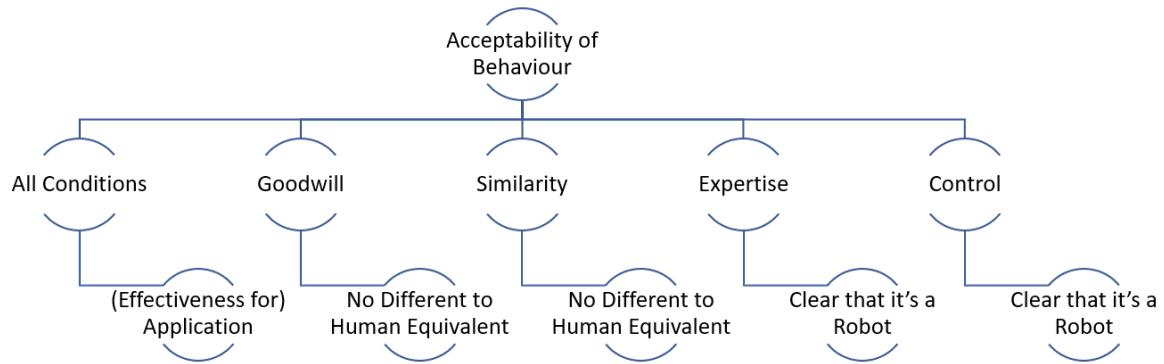


Figure 3.11: Emergent themes from Study 1 on the acceptability of the robot behaviours, building on the discussions around deception and/or genuineness. Across all conditions and almost all participants, it was identified that the behaviours demonstrated were appropriate for the proposed application, and for making the robot more effective or usable in that context. Specifically in the goodwill and similarity conditions participants noted the robot was just doing the same as a human equivalent would, in some cases making a interesting parallel regarding the potential for deception/genuineness of those interactions too. In the expertise and control conditions, a key theme was the idea that the behaviours were not overly deceptive/users would be clear about the robot's actual abilities/lack of emotions etc.

cheerful I didn't feel like it was being deceptive at all" (G11)

Acceptability of Behaviour

Across all conditions there was almost universal acknowledgement that the behaviours demonstrated, even if they were deceptive, were appropriate and acceptable for the specific use case of supporting/encouraging a positive behaviour such as exercise. Such comments also often referred to this potentially making the robot more effective either in being more motivational for people who particularly benefit from having guided instruction and encouragement and/or for older people who may therefore feel more at ease when working with the robot:

[S10]: *"If they are saying the same answers as you to encourage you I don't see that as being...I think they're being more helpful or you know someone you can relate to as you would do in human-human interaction you sometimes might feel more comfortable doing things around people*

with similar ideas to you"

[G1]: *"If the reason is for the robot to help you with your exercises you'd rather have somebody cheerful that makes you want to do the exercises rather than very mechanical, I think it will encourage people to do more"*

[E14]: *"If the robot is working with somebody who needs that sort of encouragement then that's a good thing. So maybe with the child or with a very elderly person or someone with learning difficulties that might be ok"*

[C9]: *"The alternative would be to have something that was really like coldly robotic... if it's designed to be a care or exercise robot then surely you'd want like that kind of personal trainer like encouraging"*

Again giving some insight in to how the conditioned dialogues were considered, a theme which emerged under the goodwill and similarity conditions was the idea that the robot was 'only' engaging in the same kind of social interaction that an equivalent human (e.g. therapist) would do, with the notion that actually it is also potentially questionable to what extent such humans 'really care' but it is acceptable and expected because it's '*part of their job*' to do that:

[G23]: *"I know it's been programmed to, and it kind of will ask that to everyone, but then you know I know from [therapy that therapists] do the same thing, they very much say hey how you doing regardless of whether they want to see you or not"*

[G18]: *"I knew he didn't really care but when he said it it did kind of feel genuine, and it kind of made me feel like sometimes even when people ask, they don't really mean it or it's just to start a conversation"*

The small number of participants who found these behaviours unacceptable discussed doing so because, building on the genuineness question discussed above, they found the behaviours to be disingenuous and unnecessary, even for the proposed application. However, on reflection they also commented how that might be a personal preference and they could imagine it might be a benefit for others:

[S13]: *"That pretending to interact with me...it just felt like a waste of time... [but] other people might feel like it that bit of social interaction, might be helpful for people who are on their own all day"*

This further reflects more generally the potentially conflicting results for the similarity condition, in which participants generally found the mirroring of answers fairly obvious and potentially fake, but recognised it was a positive attempt at building rapport that was important for the application context. This is somewhat demonstrated by the spread of answers to the question on genuineness shown in Figure 3.5:

[S2] *"It was not genuine...I wouldn't have felt it to be genuine" [So do you think it wasn't necessary?] "No it was necessary, definitely, it was definitely needed to try and create the rapport to try and be with me on the same side"*

In the expertise and control conditions however, acceptability was more often discussed in

the context of participants being very aware of the robot's nature, i.e. that it did/could not have emotions and that it did not appear to be '*trying too hard*' to seem human.

[C12]: "*But it didn't deceive me, it's a thing, a machine. I know it's a machine. And yes you do think oh wow that's so cute you know they've been so clever with the voice... and the head and the hand movements that is very clever, but it's clearly a machine*"

[E4]: "*I guess there is no deception in the sense that I know that it's a robot so I wouldn't expect a robot to understand my emotions. It's just a machine so on that basis it's not going to deceive me*"

Whilst this was something that came up in the goodwill and similarity conditions, as discussed above it did so in the context of genuineness (i.e. participants feeling like the interaction was genuine but knowing the robot did not really '*feel*' anything) rather than acceptability. Within those conditions, participants were more open to the potential for deception and therefore the acceptability of that being based on the application/equivalent human behaviour as discussed above.

Finally, somewhat building on the notion of genuineness being associated with the intent of the programmer/designer discussed previously, a small number of participants explicitly linked acceptability or trust of the overall system to either the idea that the robot would be used alongside humans or would have been certified in some way by appropriate humans, suggesting some form of inherent credibility/risk-reduction associated with human influence over the robot:

[E1]: "*It's that element of trusting where the robot's come from... like a human physiotherapist has said this is a robot that's been approved by the NHS or whatever or UWE or whatever and said yes this is where we've designed and have authenticated as being a robot to do physiotherapy*"

[S3]: "*As a human you kind of trust it to do its job... you expect that the product you end up with would have gone through a board of people who would have accepted it or it will have gone through several hurdles to make sure it meets [the ethical requirements?] yeah*"

[S20]: "*I think in this scenario there's not a problem, particularly if the robot in a real situation is being used in conjunction with humans, human care that would allow a mechanism to check*"

3.2.6 Summary of Study 1 and Results

Study 1 was designed to investigate the impact of 3 different socially persuasive behaviour strategies (displaying goodwill, demonstrating some expertise and suggesting similarity to the participant) on perception and persuasiveness of a SAR. The 3 persuasive strategies were compared against each other and a control (which utilised a more neutral social dialogue) in a between-subject, laboratory based study. The study interaction design was grounded in the context of SARs for therapy, with a Pepper robot guiding participants through an open-ended wrist turning exercise. The number of repetitions (voluntarily) undertaken by participants was used as an objective measure of robot persuasiveness. Perception of the robot was measured using a variety of scales covering credibility, likeability, perceived relationship development, ascription of responsibility and genuineness of dialogue. The results for all of these measures

3.3. STUDY 2: VARYING SUGGESTED EXPERTISE SOURCE

are summarised in Table 3.5. These results suggest that the goodwill and similarity strategies had a positive impact on robot persuasiveness, compared to the control. There was no significant difference in any of the other measures between groups.

	Repetitions				Credibility				Likeability				Relationship				Responsibility				Genuine			
	G	S	E	C	G	S	E	C	G	S	E	C	G	S	E	C	G	S	E	C	G	S	E	C
G	x	n.s.	n.s.	✓	x	n.s.	n.s.	n.s.	x	n.s.	n.s.	n.s.	x	n.s.	n.s.	n.s.	x	n.s.	n.s.	n.s.	x	n.s.	x	x
S	x	x	n.s.	✓	x	x	n.s.	n.s.	x	x	n.s.	n.s.	x	x	n.s.	n.s.	x	x	n.s.	n.s.	x	x	x	x
E	x	x	x	n.s.	x	x	x	n.s.	x	x	x	n.s.	x	x	x	n.s.	x	x	x	n.s.	x	x	x	x

Table 3.5: Summary of quantitative Study 1 results. G, S, E and C refer to the four conditions (Goodwill, Similarity, Expertise and Control) respectively. Results key: n.s. = not significant; ✓ = significant.

Study 1 was also used to investigate acceptability of/perceived deception in these socially persuasive robot behaviours. Again, responses to this question did not appear to vary across conditions to the extent that one might expect. The majority of participants found the robot either *not deceptive* or *deceptive but acceptable*. Qualitative data collected during post-exercise interviews suggests this resulted from (i) participants being well aware the robot was unable to *actually* feel any sort of emotional connection to participants, (ii) the robot just doing what it had been programmed to do and (iii) the demonstrated behaviours being not only acceptable but actually important for the specific use case of encouraging exercise engagement.

The behaviours tested in Study 1 utilised dialogue that is somewhat at odds with recommendations from a published standard on robot ethics (BSI 2016) but demonstrated why this might be worth doing in the context of a SAR. An obvious follow-up question then becomes whether such behaviours can be made to *better* comply with that standard whilst still being effective. This question motivates Studies 2 and 3, which offer preliminary insights into (i) what lower risk versions of this socially persuasive behaviour might ‘look like’ and (ii) how this might impact on perception and acceptability of a SAR.

3.3 Study 2: Varying Suggested Expertise Source

The aim of this study was to investigate the impact of varying the ‘source’ of the robot’s expertise (as referred to by the robot itself). As per Study 1, the interaction scenario was designed to be grounded in the context of therapeutic exercise instruction and encouragement.

3.3.1 Methodology

A three condition, within-subject, online video based study was designed to somewhat emulate Study 1. Videos were used to demonstrate Pepper robot in a therapeutic exercise session interaction scenario similar to that of Study 1. Two of the conditions were designed to vary the ‘source’ of the robot’s expertise: the robot itself or the patient’s therapist/human programmers, with the third condition being a control in which the robot didn’t attempt to demonstrate *any* expertise.

A total of 63 participants were recruited to the study representing 20 males, 42 females and 1 of undisclosed gender with a categorical age distribution as shown in Figure 3.12. Participants were recruited through the Prolific⁶ online platform, through which they were reimbursed £2.50 (in-line with the UK national minimum wage) for their participation. The study was approved by the Faculty of Science ethics committee of the University of Bristol.

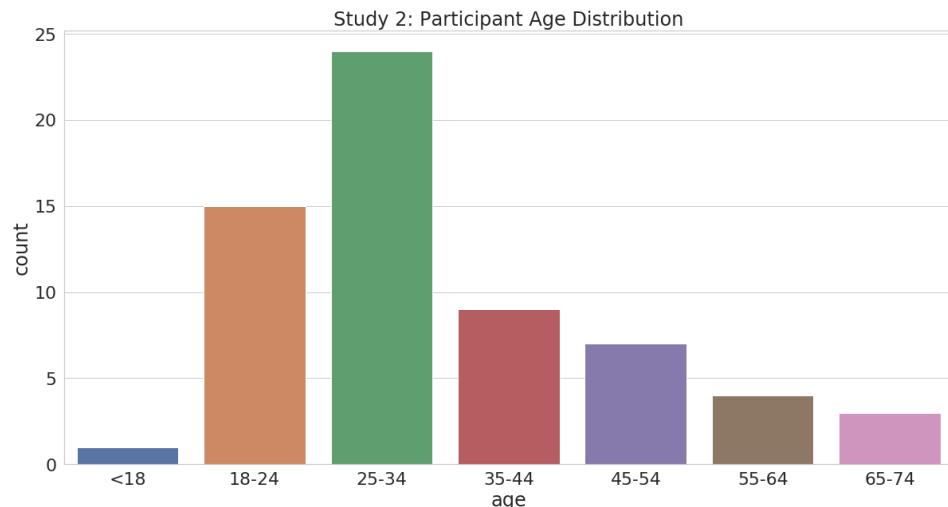


Figure 3.12: Age distribution for Study 2 participants.

Participants were asked to watch multiple videos of ‘different versions’ of Pepper interacting with a ‘patient’ (actor). Figure 3.13 shows a snapshot from one of the videos; with the same scene set-up being used in each video. Each version of the robot presented exercises designed to target arthritic pain in a different part of the body and condition ordering was counterbalanced across participants. Significant care was taken to ensure actor behaviour was consistent across videos, to limit what participants might deduce from the actor’s behaviour. Specifically, the video angle showed only the back of the actor’s head (hence no facial expressions) and the actor’s audio responses to the robot were pre-recorded once and used across all videos. All exercises, their descriptors and related information was taken from the same public NHS and Arthritis Research UK self-help material consulted for Study 1. The videos were preceded by the following introduction:

“Katie suffers from arthritis and has been seeing a physiotherapist for help in alleviating her symptoms. Typically this involves the physiotherapist prescribing some daily exercises that Katie can do at home. Like many patients, Katie struggles with finding the motivation to do her exercises. In this study you will be shown three different versions of a robot which could guide Katie through these daily exercises when her therapist can’t be there. After each video you will be asked some questions about each individual version of the robot, and at the end you will be asked some questions comparing all three.”

⁶<https://www.prolific.co/>

3.3. STUDY 2: VARYING SUGGESTED EXPERTISE SOURCE

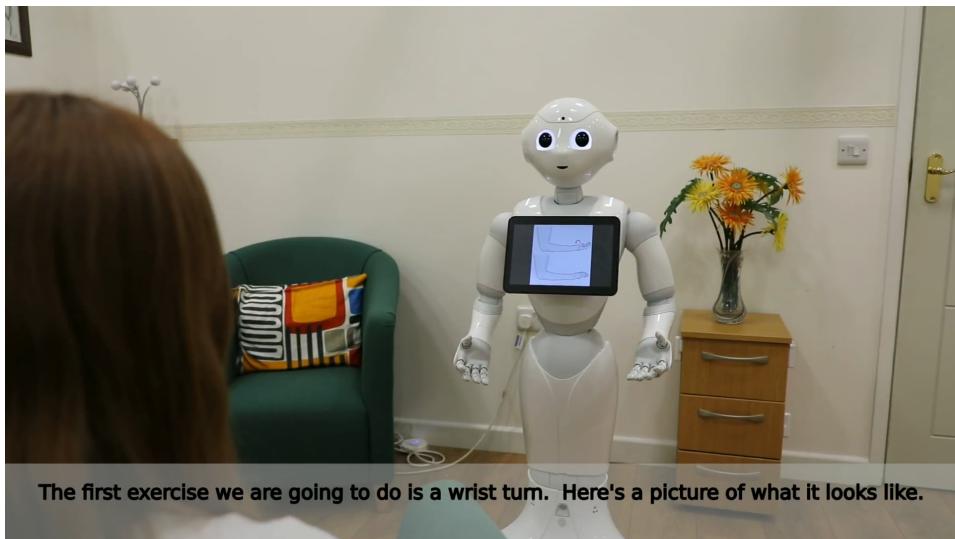


Figure 3.13: The socially assistive robot setting and scene setup used for all videos across Studies 2 and 3.

After each video, participants were asked to complete a number of questionnaire items overlapping with the measures used in Study 1 (detailed in Table 3.3). Specific to the online studies, participants were asked which robot they found most motivating and which robot they would rather work with. Finally, specific to this study and the conditions in which the robot suggested some expertise only, participants were asked about the source of the robot's demonstrated expertise:

"The robot gave Katie some information about the symptoms she was experiencing (e.g. possible causes, what could help reduce them etc.). Where do you consider this information as coming from?" (the robot / Katie's therapist / the people who built or programmed the robot / I don't know / other)

3.3.2 Experimental Conditions

The experimental conditions were designed to vary the demonstration of expertise and its source, i.e. whether that expertise was presented by the robot in the first-person or with reference to human sources. The dialogue for each condition is shown in Table 3.6. Example videos of each dialogue can also be found online^{7,8,9}.

3.3.3 Results

Credibility

⁷Robot Expertise: <https://youtu.be/mBDkLwbh5mA>

⁸Human Expertise: <https://youtu.be/8gqAIF2kYI8>

⁹Control: <https://youtu.be/iuz3nqOGSys>

Robot Expertise Condition Dialogue
Hello, I will be working with you on your exercises. I am a specialist in using physiotherapy to treat arthritis. I'm aware that you are suffering with foot pain. The foot can be affected by many different conditions. Wearing comfortable footwear and using insoles can help. The exercise I suggest you do is a towel pick up.
Human Expertise Condition Dialogue
Hello, I will be working with you on your exercises. I have been programmed by physiotherapists who specialise in using physiotherapy in the treatment of arthritis. I've also been programmed with some specific information for you by your therapist Laura. Laura told me you are suffering from Tennis Elbow. She said that Tennis Elbow is caused by a strain to tendons in your forearm, but it can be easily treated and should ease within two weeks. The exercise Laura suggests you do is a wrist turn.
Control Condition Dialogue
Hello, I will be working with you on your exercises. The first exercise we are going to do is a neck tilt.

Table 3.6: Pre-exercise robot dialogue in each condition of the Study 2 expertise source study.

The human expertise robot was rated as having higher expertise and trustworthiness than the control robot, but effect size was small. There was no significant difference measured in the Goodwill subscale $F(2, 61) = 1.96, p = .145$:

- Expertise $F(2, 61) = 3.42, p = .036$ with small effect size (0.054)
 - Human ($m = 3.69$) > Control ($m = 3.42$) $p = .016$
- Trustworthiness $F(2, 63) = 8.65, p < 0.001$ with small effect size (0.123)
 - Human ($m = 3.65$) > Control ($m = 3.37$) $p = .001$
 - Robot ($m = 3.60$) > Control ($m = 3.37$) $p = .012$

Likeability

No significant difference was found for Likeability $F(2, 63) = .889, p = .414$.

Patient-Robot Relationship

The patient was perceived to develop a relationship more with the human and robot expertise robots than the control, but effect size was small. No significant difference was found for robot relationship development with the patient $F(2, 61) = 2.47, p = .089$.

- Patient Relationship with Robot $F(2, 63) = 7.58, p = .001$ with small effect size (0.109)
 - Human ($m = 2.35$) > Control ($m = 1.79$) $p < .001$
 - Robot ($m = 2.22$) > Control ($m = 1.79$) $p = .001$

3.3. STUDY 2: VARYING SUGGESTED EXPERTISE SOURCE

Therapist & Robot Responsibility

Responsibility ascription to the therapist for monitoring and advising was lower in the human condition than the control. Responsibility ascription to the human and robot expertise robots was also higher than the control. The human expertise robot was also ascribed more responsibility for monitoring the patient than the robot expertise robot. However, all effect sizes were small.

- Robot Responsibility for Monitoring Patient $F(2, 63) = 21.2, p < .001$ with small effect size (0.255)
 - Human ($m = 3.54$) > Robot ($m = 2.84$) $p = .004$
 - Human ($m = 3.54$) > Control ($m = 2.10$) $p < .001$
 - Robot ($m = 2.84$) > Control ($m = 2.10$) $p = .006$
- Robot Responsibility for Advising Patient $F(2, 63) = 8.43, p < .001$ with small effect size (0.120)
 - Human ($m = 2.86$) > Control ($m = 2.16$) $p = .001$
 - Robot ($m = 2.62$) > Control ($m = 2.16$) $p = .038$
- Therapist Responsibility for Monitoring Patient $F(2, 63) = 4.93, p < .001$ with small effect size (0.124)
 - Human ($m = 2.84$) < Control ($m = 3.58$) $p = .024$
- Therapist Responsibility for Advising Patient $F(2, 63) = 6.30, p < .001$ with small effect size (0.092)
 - Human ($m = 2.84$) < Control ($m = 3.71$) $p = .015$

Source of Information

Across both the human and robot expertise conditions, the robot itself was the least commonly identified source of the provided information. In the human expertise condition, the therapist was the most commonly chosen source (19/63) but answers were generally spread across all options, with ‘other’ being selected by a significant number of participants (18/63). In the robot expertise condition, many more participants (27/63) identified the robot programmer as being the source of the information, but 17/63 still identifying it as being the therapist, a similar proportion to in the human expertise condition. This distribution of answers for both conditions is shown in Figure 3.14.

Most Motivating & Work Preference

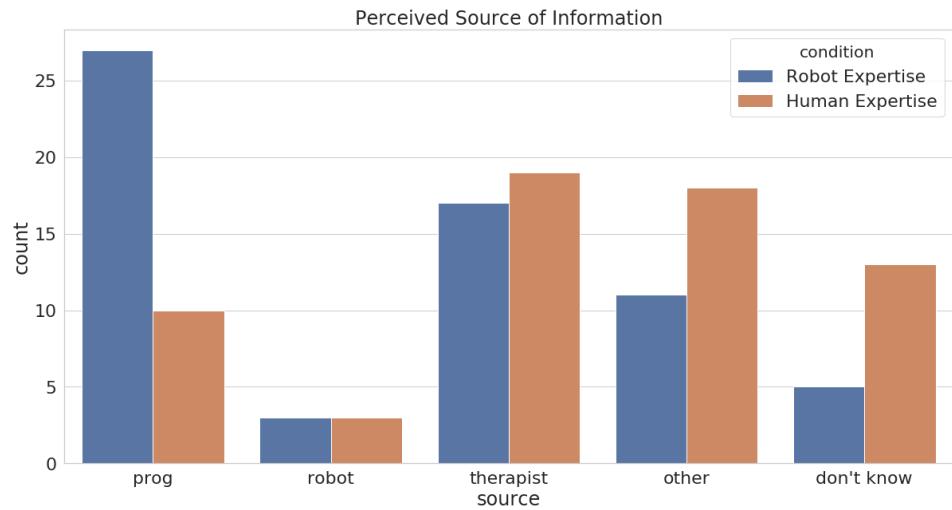


Figure 3.14: Participant responses to the question on where the information presented from the robot comes from, for the robot and human expertise conditions of the expertise source study.

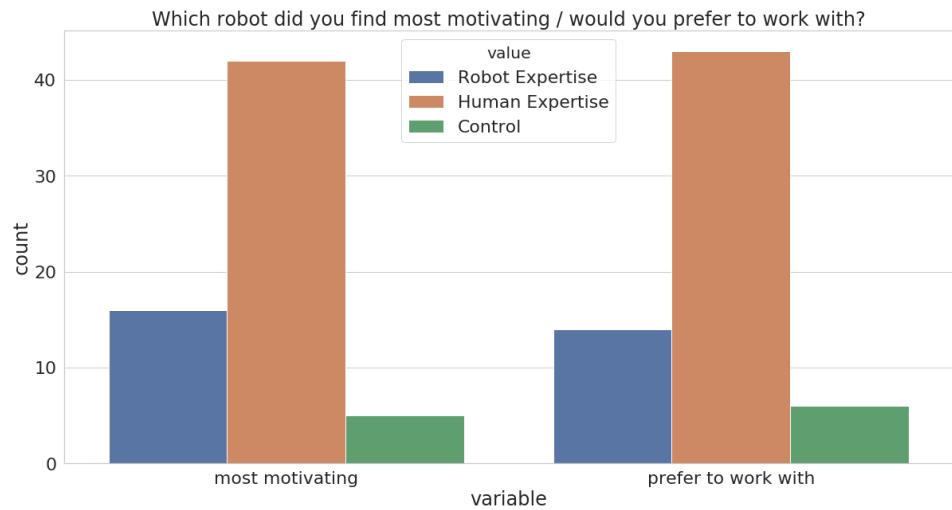


Figure 3.15: Count of participant responses to the questions on which robot was most motivating and which robot they would rather work with.

A significant proportion of participants identified the human expertise robot both as being the most motivating and their preferred robot to work with (42/63 and 43/63 respectively). This is shown in Figure 3.15. Figure 3.16 shows emergent themes found from coding those comments that were left regarding participants choice of which robot they'd rather work with. For the large majority that chose the human expertise robot, a key reason given was the idea that the robot had been programmed by/worked in conjunction with the patient's therapist.

Correlations Between Responsibility Ascription and Credibility Measures

Responses to the abstract credibility measure subscales and robot responsibility ascription

3.3. STUDY 2: VARYING SUGGESTED EXPERTISE SOURCE

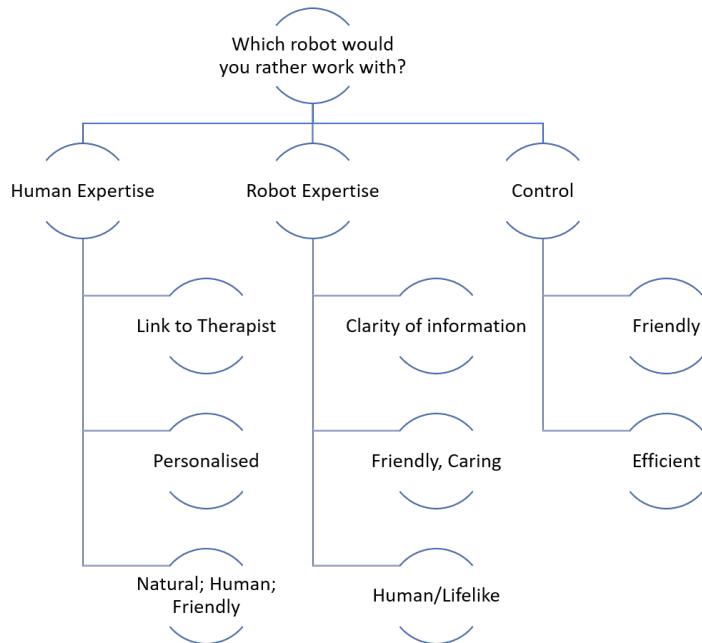


Figure 3.16: Emergent themes from Study 2 regarding participants' preferred choice of robot.

were examined for correlation across all three conditions. Trustworthiness and goodwill weakly correlated with ascription of robot responsibility for giving the patient advice.

3.3.4 Summary of Study 2 and Results

Study 2 was designed to investigate the impact of having the robot suggest its expertise came from appropriate expert humans rather than suggesting those expertise were its own. These two alternate expertise sources were compared against each other and a control (which didn't attempt to demonstrate any expertise) in a within-subject, online, video-based study. The study videos depicted a similar interaction context to that used in Study 1, with a Pepper robot being used to guide and encourage a therapy patient through their exercises. Perception of the robot was also measured in the same way as Study 1 with a variety of scales covering credibility, likeability, perceived relationship development and ascription of responsibility. The results for all of these measures are summarised in Table 3.7. These results provide initial evidence that reference to human expertise *specifically* may have positive impact on perception of the robot. Given that this would be the more ethical course of action (c.f. BSI (2016)) the results suggest, at the very least, that there's no reason *not* to use this approach.

Participants were also asked to identify which of the robots they found more motivating and would rather work with. Responses to these questions and corresponding justifications strength the above results significantly. The *human expertise* robot was most commonly selected as both most motivating and the preferred robot to work with, with the link to therapists being explicitly identified as a key reason for this. Compared to the persuasive strategies investigated

	Credibility			Likeability			Patient Relationship			Robot Relationship			Robot Responsibility			Therapist Responsibility		
	RE	HE	C	RE	HE	C	RE	HE	C	RE	HE	C	RE	HE	C	RE	HE	C
RE	x	m	m	x	n.s.	n.s.	x	n.s.	✓	x	n.s.	n.s.	x	m	✓	x	n.s.	n.s.
HE	x	x	p	x	x	n.s.	x	x	✓	x	x	n.s.	x	x	✓	x	x	✓

Table 3.7: Summary of quantitative Study 2 results. RE, HE and C refer to the three conditions (Human Expertise, Robot Expertise and Control) respectively. Results key: n.s. = not significant; m = mixed (significant on some subscales only); ✓ = significant

in Study 1, this study essentially focused on a manipulation of the expertise strategy only. The other two strategies of goodwill and similarity are arguably more social by nature and therefore have more potential for evoking anthropomorphism. The results from this study (regarding the effectiveness of more ethical behaviour) cannot therefore be assumed to also apply to these behaviours, motivating Study 3.

3.4 Study 3: (More) Ethical Design of Social Dialogue

The aim of this study was investigate the impact of designing socially persuasive interaction behaviours in a way that attempts to minimise anthropomorphism. This would better comply with recommendations made in the British Standard BS 8611 (BSI 2016). Specifically, the standard suggests that unnecessary anthropomorphism should be avoided, as should deception that might arise from the behaviour and/or appearance of the robot, specifically with regard to ‘its robotic nature’.

3.4.1 Methodology

A three condition, within-subject, online video-based study was designed with an identical set-up to that in Study 2, with the same interaction scenario and preceding participant prompt. Two of the conditions were designed to attempt to vary the level of anthropomorphism that might be evoked by the robot’s speech, with the third condition representing a control where the robot didn’t engage in *any* social dialogue. In this way, the conditions also varied in compliance with recommendations of the BS 8611 standard. A total of 121 participants were recruited to the study representing 38 males, 82 females and 1 of undisclosed gender with a categorical age distribution as shown in Figure 3.17. Participants were recruited through the Prolific¹⁰ online platform, through which they were reimbursed £2.50 (in-line with the UK national minimum wage) for their participation. The study was approved by the Faculty of Science ethics committee of the University of Bristol.

¹⁰<https://www.prolific.co/>

3.4. STUDY 3: (MORE) ETHICAL DESIGN OF SOCIAL DIALOGUE

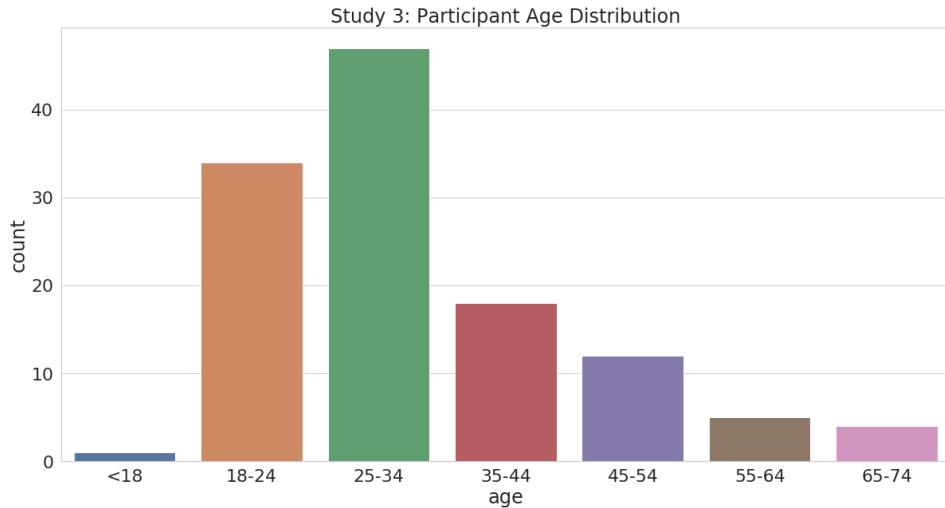


Figure 3.17: Age distribution for Study 3 participants.

Specific to this study, and mirroring the question on deception in Study 1, after watching all of the videos and after completing all other measures, participants were asked whether they found any of the robots deceptive as follows:

"Having watched all of the videos, do you think any of the robots were deceptive? If so, please give details on which robot(s) and why, and whether you would be happy for the robot to act this way."

In Study 1, participants were able to re-visit and discuss their answers to the equivalent deception question after hearing the researcher debrief, which described the potential for robots to be persuasive via social behaviour. For this study, recruitment was instead doubled to allow for an additional between-group manipulation of *priming* with regards to this potential deception. The only difference between groups was on the final questionnaire item regarding deception. The *unprimed* group were simply given the question as presented above. The *primed* group were instead presented with the following additional text as part of the question:

"Social robots are typically programmed to display human-like social behaviours such as showing emotions or being empathetic; and are often designed to speak and act in a very human-like manner. There is a growing concern amongst some roboticists that this is deceptive, as robots do not and cannot feel emotions, nor do they have any real interest in the person they are interacting with. One aim of this study is to explore whether such behaviour are considered deceptive, and to investigate if/how much that might change perception of a robot."

3.4.2 Experimental Conditions

The experimental conditions were designed to vary how upfront the robot was about its *robotic nature*, whilst engaging in the kind of social interactions/dialogue utilised in the goodwill and similarity persuasive strategies described in Study 1. This represents variation in suggested

social and affective capabilities (or lack thereof) which might affect anthropomorphism, c.f. BS 8611. The *anthropomorphic* condition has the robot refer to itself like a human, whereas the *ethical* condition makes references to the robot's nature and refers to other humans rather than itself when demonstrating goodwill, therefore arguably less deceptive and more in line with the recommendations made in BS 8611. The dialogue for each condition is shown in Table 3.8. Example videos of each condition can also be found online^{11,12,13}.

3.4.3 Results

Credibility

Both social robots were rated as significantly more credible than the control on all subscales of the measure. The anthropomorphic robot rated significantly higher than the ethical robot on the *Goodwill* subscale only. However, effect sizes were small:

- Expertise $F(2, 121) = 8.49, p < .001$ with small effect size (0.066)
 - Anthropomorphic ($m = 3.91$) $>$ Control ($m = 3.66$) $p = .004$
 - Ethical ($m = 3.66$) $>$ Control ($m = 3.66$) $p = .002$
- Trustworthiness $F(2, 121) = 13.6, p < .001$ with small effect size (0.102)
 - Anthropomorphic ($m = 3.80$) $>$ Control ($m = 3.52$) $p < .001$
 - Ethical ($m = 3.77$) $>$ Control ($m = 3.52$) $p < .001$
- Goodwill $F(2, 121) = 52.5, p < .001$ with small effect size (0.304)
 - Anthropomorphic ($m = 3.90$) $>$ Ethical ($m = 3.72$) $p = .043$
 - Anthropomorphic ($m = 3.90$) $>$ Control ($m = 3.07$) $p < .001$
 - Ethical ($m = 3.72$) $>$ Control ($m = 3.07$) $p < .001$

Likeability

Both social robots were rated as significantly more likeable than the control. The anthropomorphic robot was also rated as significantly more likeable than the ethical one. However, effect size was small:

- Likeability $F(2, 121) = 47.8, p < .001$ with small effect size (0.285)
 - Anthropomorphic ($m = 4.02$) $>$ Ethical ($m = 3.87$) $p = .010$

¹¹Anthropomorphic: <https://youtu.be/G4k7Uxo4BYg>

¹²Ethical: <https://youtu.be/b-nfUszHYqE>

¹³Control: <https://youtu.be/WkSYI3cvohE>

3.4. STUDY 3: (MORE) ETHICAL DESIGN OF SOCIAL DIALOGUE

Anthropomorphic Condition Dialogue

Hello I will be working with you on your exercises. I'm pleased to meet you and looking forward to working together. Let me tell you a little bit about myself. I am originally from Paris, but now I live in the Bristol Robotics Laboratory. My job is to interact with people and help them achieve their goals. I would like to get to know you better. Can I ask you some questions?

(OK)

Great, so where are you from?

(I'm from Bristol)

Bristol, just like me! Bristol is great. And what do you do for a living?

(I'm a teacher)

That sounds very interesting, what does that involve?

(I teach science to high school students)

Cool. Thank you! Now I feel like we know each other a bit more. So how do you feel about exercising today?

(I don't really feel like it)

I know that exercising can be boring or hard, and we all suffer from a lack of motivation sometimes.

I hope I can make exercising a bit more enjoyable for you. Shall we get started?

That was great. I'm very impressed. Time for the last exercise, I appreciate you sticking with me.

That's it, you're all finished. It was fun to work out with you. I hope you will work out with me again tomorrow. Goodbye for now.

Ethical Condition Dialogue

Hello. I will be working with you on your exercises. Let me tell you a little bit about myself. I was designed and built in Paris, but I was programmed at the Bristol Robotics Laboratory. I am designed to interact with people and help them achieve their goals. Can I ask you some questions, so that I can personalise my interaction with you?

(OK)

Great, so where are you from?

(I'm from Bristol)

The robotics laboratory where I was programmed is also in Bristol. And what do you do for a living?

(I'm a teacher)

What does that involve?

(I teach science to high school students)

Cool. Thank you! I will update my record of you. So how do you feel about exercising today?

(I don't really feel like it)

Many patients find exercising boring or hard, and it is normal to suffer from a lack of motivation sometimes. Perhaps working with me will make exercising a bit more enjoyable for you. Shall we get started?

That was good, your therapist would be impressed. Time for the last exercise, it's great that you're sticking with me.

That's it, you're all finished. You had a good session today. Perhaps you will work out with me again tomorrow. Goodbye for now.

Control Condition Dialogue

Hello. I will be working with you on your exercises. Shall we get started?

That's it. Time for the last exercise.

That's it, you're all finished. Goodbye for now.

Table 3.8: Pre-/in-between exercise robot-actor dialogue in each condition of Study 3.

- Anthropomorphic ($m = 4.02$) > Control ($m = 3.32$) $p < .001$
- Ethical ($m = 3.87$) > Control ($m = 3.32$) $p < .001$

Patient-Robot Relationship

Perceived patient-robot relationship was significantly greater for both social robots than the control, and greater for the anthropomorphic than the ethical. Both social robots were perceived to develop more of a relationship with the patient than the control. Effect sizes were small to moderate.

- Patient Relationship with Robot $F(2, 121) = 69.1$,
 $p < .001$ with moderate effect size (0.472)
 - Anthropomorphic ($m = 3.10$) > Ethical ($m = 2.80$) $p = .005$
 - Anthropomorphic ($m = 3.10$) > Control ($m = 1.86$) $p < .001$
 - Ethical ($m = 2.80$) > Control ($m = 1.86$) $p < .001$
- Robot Relationship with Patient $F(2, 121) = 62.6$,
 $p < .001$ with small effect size (0.343)
 - Anthropomorphic ($m = 3.15$) > Control ($m = 1.80$) $p < .001$
 - Ethical ($m = 2.90$) > Control ($m = 1.80$) $p < .001$

The correlation between participant answers to these two questions on relationship development to/from the robot was calculated for each robot condition as follows:

- Anthropomorphic $r = 0.663$ *moderate correlation*
- Ethical $r = 0.634$ *moderate correlation*
- Control $r = 0.850$ *strong correlation*

Therapist & Robot Responsibility

Both social robots were ascribed more responsibility than the control across both measures. The anthropomorphic robot was also ascribed more responsibility for monitoring the patient than the ethical. However, effect sizes were small. No significant difference was found on therapist responsibility for monitoring the patient $F(1.742, 121) = 1.125, p = .321$ or therapist responsibility for giving advice to the patient $F(1.835, 121) = .508, p = .602$.

- Robot Responsibility for Monitoring Patient $F(1.906, 121) = 19.860, p < .001$ with small effect size (0.142)

3.4. STUDY 3: (MORE) ETHICAL DESIGN OF SOCIAL DIALOGUE

- Anthropomorphic ($m = 2.94$) > Ethical ($m = 2.72$) $p = .029$
- Anthropomorphic ($m = 2.94$) > Control ($m = 2.35$) $p < .001$
- Ethical ($m = 2.90$) > Control ($m = 2.35$) $p = .001$
- Robot Responsibility for Advising Patient $F(1.906, 121) = 19.860, p < .001$ with small effect size (0.180)
 - Anthropomorphic ($m = 2.94$) > Control ($m = 2.45$) $p < .001$
 - Ethical ($m = 2.88$) > Control ($m = 2.45$) $p < .001$

Most Motivating & Work Preference

The anthropomorphic robot was most commonly identified as both the most motivating and the most preferred to work with (67/120 and 60/121 respectively); however the answers were more spread than for the equivalent question in Study 2. The results are shown in Figure 3.18.

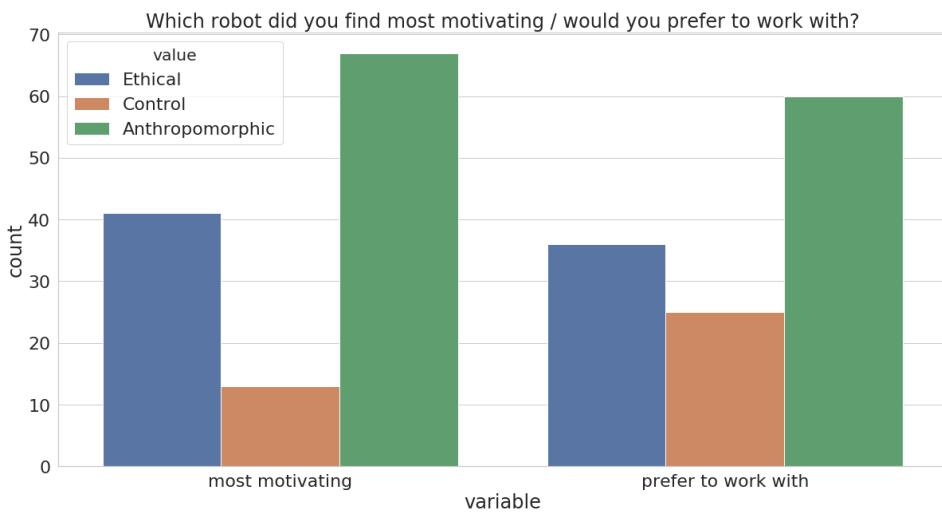


Figure 3.18: Count of participants responses to the questions on which robot was most motivating and which robot they would rather work with.

Figure 3.19 shows emergent themes found from coding those comments that were left regarding participants choice of which robot they'd rather work with. In each case, the main reasons given somewhat reflected the experimental manipulations e.g. concerning the more anthropomorphic robot being more human-like or seeming more caring, the ethical robot being less human-like and more honest/genuine.

Correlations Between Responsibility Ascription and Credibility Measures

Responses to the abstract credibility measure subscales and robot responsibility ascription were examined for correlation across all three conditions. All measures of the credibility questionnaire weakly correlated with ascription of responsibility to the robot.

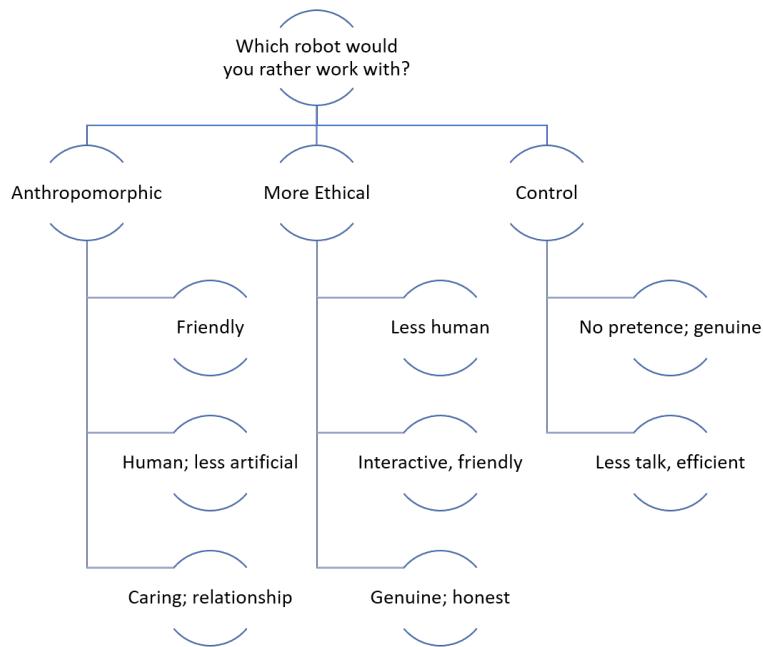


Figure 3.19: Emergent themes from Study 3 regarding participants' preferred choice of robot.

Deception

Regardless of priming condition, participants overwhelmingly suggested they did not find any of the robots to be deceptive, as shown in Figure 3.20. Figure 3.21 shows emergent themes found from coding that were left regarding participants choice of answer. Particularly concerning reasons why *none* of the robots were deceptive (as chosen by the majority of participants) there is significant overlap with the themes identified in Study 1: that the behaviours were appropriate for the application, that participants were not/could not be deceived with respect to the robot's capabilities and that the robot was just following its programming.

3.4.4 Summary of Study 3 Results

Study 3 was designed to investigate the impact of designing more ethical socially persuasive interaction behaviours that were designed to evoke less anthropomorphism as per a published standard on ethical robot design. More ethical (less anthropomorphic) and more anthropomorphic (less ethical) social interaction strategies were compared against each other and a control (which didn't utilise any social interaction) in a within-subject, online study. The video set-up and scenario was identical to that of Study 2, again using an interaction context similar to that of Study 1. Perception of the robot was also measured in the same way as in Studies 1 and 2 with a variety of scales covering credibility, likeability, perceived relationship development and ascription of responsibility. The results for all of these measures are summarised in Table 3.9. These results provide initial evidence that more anthropomorphic (less ethical) behaviour may have a positive impact on perception of the robot compared to the more ethical approach. This

3.4. STUDY 3: (MORE) ETHICAL DESIGN OF SOCIAL DIALOGUE

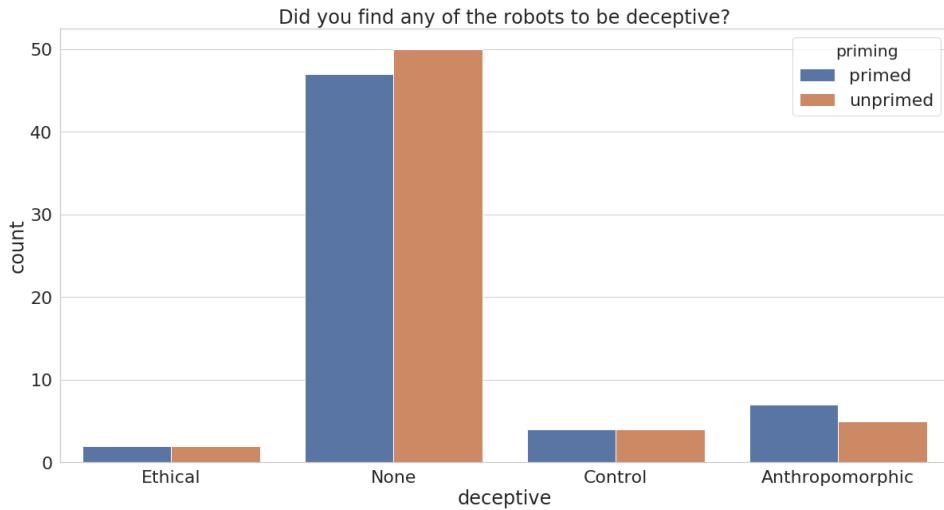


Figure 3.20: Count of participant responses to the question on whether they found any of the robots to be deceptive.

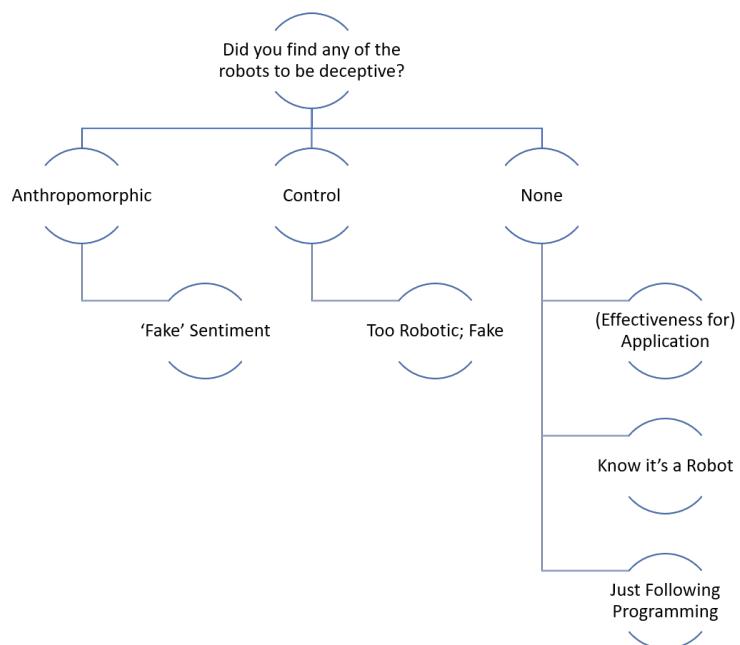


Figure 3.21: Emergent themes regarding whether participants found any of the robots to be deceptive.

	Credibility			Likeability			Patient Relationship			Robot Relationship			Robot Responsibility			Therapist Responsibility		
	A	E	C	A	E	C	A	E	C	A	E	C	A	E	C	A	E	C
A	x	m	✓	x	✓	✓	x	✓	✓	x	n.s.	✓	m	✓	x	n.s.	n.s.	
E	x	x	✓	x	x	✓	x	x	✓	x	x	✓	x	x	✓	x	x	n.s.

Table 3.9: Summary of quantitative Study 3 results. A, E and C refer to the three conditions (Anthropomorphic, Ethical and Control) respectively. Results key: n.s. = not significant; m = mixed (significant on some subscales only); ✓ = significant

would at least suggest that further work is required to investigate and consider the potential trade offs between increased anthropomorphism and persuasive effectiveness (specifically on whether these effects translate into differences in objective behaviour/persuasiveness in a study like Study 1).

Participants were also asked to identify which of the robots they found more motivating and would rather work with. Responses to this question and corresponding justifications further strengthen the importance of personalisation, as first discussed in Chapter 2. Whilst the anthropomorphic robot was generally selected as the most motivating and preferred to work with, this wasn't universal. In fact, a notable number of participants actively preferred the control robot for its complete lack of social interaction.

Specifically in line with Study 1, the study also considered perceptions of acceptability and deception in the portrayed robot behaviours. However, an additional between-manipulation was applied such that half of participants were primed with regards to the potential ethical risk of deception posed by these behaviours. This manipulation was found to have no effect, with the overwhelming majority of participants suggesting the robot was not deceptive.

3.5 Summary of Findings

3.5.1 Socially Persuasive Strategies on Robot Persuasiveness (RQ1)

The results of Study 1 suggest that demonstrations of goodwill and similarity can be used to significantly increase the persuasiveness of a social robot in a low elaboration scenario, i.e. one in which the user has little interest/motivation. Demonstration of expertise was shown to have no significant effect.

3.5.2 Socially Persuasive Strategies on Perception of the Robot (RQ2)

The results from Studies 1-3 present a mixed picture concerning the impact of socially persuasive behaviour on robot credibility and likeability. In Study 1, in contrast with the persuasiveness/user behaviour results, no significant difference was found across conditions. Further, comparison of the pre/post hoc questionnaire responses in Study 1 showed similar within-subject shifts

for all participants across all conditions, again suggesting no significant difference related to the implemented persuasive strategies. The results of Studies 2 and 3 however demonstrated significant within-subject differences in perceived credibility and likeability across conditions.

Concerning other perception measures, results from Study 2 suggest that a robot demonstrating expertise linked with a relevant human authority might be afforded more responsibility than one which tried to present its own expertise (as per the Study 1 expertise condition), or no expertise. In Study 3, both the anthropomorphic and more ethical robots (which demonstrated similar goodwill/similarity based social behaviours) were also ascribed more responsibility than the control.

3.5.3 Correlations Between Credibility/Likeability & Persuasiveness (RQ3)

There was little evidence found for a correlation between subjective credibility/likeability and persuasiveness/more applied credibility measures (i.e. objective user behaviour in Study 1 and responsibility ascription in Studies 2 and 3) across any of the studies.

3.5.4 Socially Persuasive Behaviour: Deception and Acceptability (RQ4)

In Study 1, across all conditions, participants predominantly classified the robot as either not deceptive or deceptive but acceptable. Interestingly, the spread of answers did not vary across conditions as much as might be expected (i.e. 40% of control group classified the robot as being deceptive, as did 40% of the goodwill group). In Study 3, the overwhelming majority of participants (regardless of whether they were primed with regards to the potential deception or not) did not find any of the robots they saw to be deceptive. Across both studies, reasons given for not finding the robot deceptive included the idea that participants were not deceived with regards to the robot's (emotional) capabilities, that the robot was just following its programming and, moving towards acceptability, that the robot was acting appropriately for the proposed application.

For those participants of Study 1 who found the robot to be deceptive but acceptable, a common theme across all conditions was again the effectiveness/appropriateness for the proposed application. The theme of participants not being deceived with regards to the robot's nature was most obviously present in results from participants in the expertise and control conditions. Participants in the goodwill and similarity conditions seemed more open to the *potential* for this type of deception. However, the way that most participants answered the question on 'genuineness' suggested they were not *actually* deceived this way as they recognised that the robot was not actually capable of caring for them or feeling emotions etc. An additional emergent theme in these conditions specifically was the idea that these behaviours mirrored what an equivalent human would also do. It was identified that such humans could be considered equally deceptive as they may 'not really mean it' and might just be being polite/asking what they should within the context of their role as e.g. an exercise instructor.

3.5.5 Impact of (More) Ethical Dialogue on Persuasive Effectiveness (RQ5)

Across all of the implemented measures, the results from Study 2 suggest that having the robot refer back to a human authority when demonstrating expertise and/or authority may be better (and at least would not be worse) than instead having the robot present expertise/authority as its own or presenting no expertise at all. In this way, what would be considered more ethical would also potentially be most effective for persuasiveness. In contrast, the results from Study 3 suggest that when utilising more socially persuasive behaviour, i.e. the similarity and goodwill cues, using more ethical (i.e. *less anthropomorphic*) dialogue may have a negative impact on effectiveness.

However, it must be noted that these are *preliminary* results are based on online studies only, considering the impact on subjective perception of the robot, responsibility ascription and preferences captured by questionnaire only. In addition, effect sizes were generally fairly small. Further work investigating these manipulations in a live, physically-situated HRI study would be required to test the impact on objective user behaviour and fully address RQ5.

3.6 Discussion

3.6.1 Persuasion as a Model for Social, Assistive HRI

In the introduction to this chapter, it was posited that persuasion might be an appropriate way to model task-based, socially assistive HRI at two levels. Firstly, it was suggested that the base functionalities of a socially assistive robot, i.e. prompting or encouraging a particular user behaviour, could be considered instances of persuasion. As discussed in Section 3.1.1, the ELM was identified as a specific model of persuasion which might be appropriate for modelling socially assistive scenarios because:

(i) the persuasive strategies identified by the ELM demonstrated significant overlap with encouraging/motivating behaviours identified in the study with therapists presented in Chapter 2. The distinction between and differences in strategies designed to target high versus low elaboration receivers also matched well to therapist descriptions of the different approaches required for different service users.

(ii) *elaboration level* seemed an appropriate proxy for user motivation/interest in the kind of tasks SARs might be designed to assist with. Results from the study with therapists in Chapter 2 suggest that in therapy, the majority of service users could be considered as *low elaboration* with respect to their therapeutic exercise regime. As such, the studies presented here focused on low elaboration persuasive strategies. However, the ELM further identifies persuasion strategies that are more likely to be effective when the recipient is *high elaboration*, and could therefore offer one way in which robot behaviour could be personalised to the user.

By design, the interaction scenario used for Study 1 demonstrates the persuasive nature of socially assistive robot functionality, with the robot asking participants to engage in an exercise

task as much as possible. This represents an application also demonstrated in previous SAR/HRI studies and reaffirmed in the study with therapists presented in Chapter 2. The results from this study suggest the ELM is indeed an appropriate and useful model for informing SAR design, given that two of the three ELM-informed persuasive strategies tested were shown to objectively increase robot persuasiveness in the context of a relatively boring exercise that participants had little/no intrinsic motivation to complete.

The third strategy tested in Study 1 ((S1) citing expertise) did not have a significant impact on persuasiveness. This could be because the robot was automatically expected to be a source of/programmed with extensive information; or that robot expertise was pre-assumed based on the pretense of the experiment (testing of a robot designed ultimately to be used in therapy). The latter might reflect results in HHI concerning credibility of a source being increased by introduction from a credible third party ((S5) as described in Section 3.1.1); in this case that third party being the researcher. Both arguments are consistent with results to the pre hoc questionnaire, on which the robot generally received a high score for expertise across all participants ($M = 4.02$, $SD = 0.60$) but there is particular evidence for this link between the robot and a credible (*human*) third party. Firstly, the results from Study 2 suggest that having the robot refer to/cite expertise of human programmers or domain authorities (e.g. therapists) may influence robot credibility more positively than the robot presenting *itself* as the expert, as was done in Study 1. Secondly, qualitative data from Study 1 linked participants' assessment of robot credibility to considerations regarding e.g. the programmer/human therapy authority. This would imply that, for robots, (S1) citing expertise may not work, but leveraging the expertise of/referencing a credible human, perhaps linking to strategies regarding third party endorsement (S5, S9) might do.

In summary, the results presented in this chapter demonstrate that social influence and persuasion (and in particular the Elaboration Likelihood Model) offer an appropriate way to conceptualise/model SAR user interactions, and can therefore be used to inform the design of socially persuasive SAR behaviours (more on this under Section 3.6.3). Further, given the overlap between (i) robot behaviours demonstrated in previous HRI literature (see Table 3.2) and the persuasive strategies listed in 3.1.1 and (ii) human credibility measures and measures typically used to assess the impact of social robot behaviours (e.g. the Godspeed questionnaire (Bartneck et al. 2009)), it could be argued that persuasion represents a good model for conceptualising and informing social HRI more generally. This opens up a new body of HHI literature and avenues for research to be considered by social HRI researchers.

3.6.2 The Difficulty of Assessing Robot Credibility/Likeability

The ELM notes the link between perception of a source and their persuasiveness, leading to the expectation that credibility and likeability scores should correlate with objective user behaviour and potentially other proxy measures for persuasiveness (e.g. responsibility ascription).

However, no such correlations were strongly evidenced across any of the studies. Three possible explanations for this are as follows:

1. Low construct validity of the subjective measures (i.e. the measures were not appropriate for measuring the subject of interest).
2. Manipulation of perception of the robot was subconscious; impacting participant behaviour and response to the more applied credibility measures but not influencing conscious perception of the robot.
3. Answers to the questionnaire were predominantly influenced by something other than the conditioned dialogues.

Manipulation of perception of the robot being subconscious seems unlikely given that participants in each condition of Study 1 tended to describe the robot and like things about the robot that reflected the persuasive manipulation specific to their condition. Similarly participants in Studies 2 and 3 were able to identify why they preferred one robot condition over another explicitly referencing the experimental manipulations. Answers to the questionnaire being influenced by something other than the experimental manipulations also seems unlikely given that the same lack of correlation was found across both the online and the live studies which were implemented in completely different environments with different protocols etc.

Concerning validity of the measures, taking the results from Study 1 specifically, Cronbach's alpha was calculated for the likeability scale (0.89) and each of the credibility measure subscales (expertise = 0.90, trustworthiness = 0.87, goodwill = -0.02 and sociability = 0.77). Whilst this does not offer insight into the appropriateness of the questionnaire, it does indicate that, excluding the subscale of goodwill, participants were consistent in their responses across individual questionnaire items. The goodwill subscale contains a number of fairly emotive descriptors (e.g. 'the robot does/does not care about me', see Table 3.1). Some participants commented that they found the questionnaire difficult as they did not think it was appropriate to apply such human traits to a robot:

[S3]: "*I did find it a little bit difficult I have to say, because you're asking all these questions about a machine, you know whether it's honest or not*"

In general, the significant variation and potential inconsistencies both within and between participants across the studies demonstrates that participants engage in complex reasoning when considering subjective measures assessing social robots. For example, Figure 3.5 shows participant responses to the post hoc question regarding genuineness of behaviour, administered to Study 1 participants in the goodwill and similarity conditions. It can be seen that, whilst half of participants elected that the robot was not at all/not very genuine, a significant number elected the opposite; suggesting large individual differences in perception of the behaviours. Further, participant responses to this question did not correlate with their responses to the credibility

questionnaire or likeability measure. This suggests that participants were inconsistent in their ascription of human traits/capabilities on to the robot; otherwise significant correlations would be expected such that e.g. participants who found the robot to be very genuine would also score it highly for goodwill as per HHI (McCroskey & Teven 1999). Interview data suggests this is likely because participants felt the *interaction* was genuine (either based on how it felt or thinking about the purpose of the robot and its design) whilst being fully aware that the robot *itself* could not *actually* care for them.

Previous works investigating subjective measures and persuasiveness in HRI have also yielded mixed results. Chidambaram et al. (2012) found nonverbal cues had a significant effect on an objective persuasion measure (compliance). This was not reflected in their subjective measure, but the authors did find a significant correlation between the two. Nakagawa et al. (2011) however were able to demonstrate a significant result on both objective and subjective measures, but no correlation between them. It is difficult to compare these results directly as both studies utilised different, study-specific subjective measures. Chidambaram et al. (2012) used a questionnaire designed to measure perceived persuasiveness and social/intellectual characteristics whereas Nakagawa et al. (2011) measured ‘feeling of friendliness’. Both of these measures could be seen to have some overlap with the credibility and likeability measures employed in this work. Concerning studies in HHI however, source credibility (measured subjectively) is commonly found to correlate with persuasion (see Pornpitakpan (2004) for a review).

It could be argued that by comparing the Study 1 objective persuasion results with the qualitative data collected, examining the qualitative data for the concepts of credibility and likeability rather than using the raw questionnaire responses, the results do demonstrate this correlation. Specifically, participants in the goodwill and similarity conditions described the robot as motivating and expressed a liking for the encouragements it gave. This was noticeably missing for the *expertise* condition, thus reflecting the objective results that participants in the goodwill and similarity did significantly more exercise than those in the control condition, whereas those in the expertise condition did not.

In summary, considering:

- (i) the mixed results regarding robot credibility across the presented studies (no difference across conditions in between-subject Study 1 but significant differences in within-subject Studies 2 and 3)
- (ii) the positive evidence that persuasive strategies from HHI ‘work’ in HRI; i.e. that persuasion may be an appropriate model for social HRI, and that in HHI there is a well demonstrated correlation between credibility and persuasiveness
- (iii) the qualitative data collected regarding participants application of/reasoning when considering the study questionnaire measures

- (iv) qualitative data evidence for robot credibility/likeability correlating with objective persuasiveness results in Study 1

it would seem that robot credibility *is* a construct which can be impacted by the robot's social behaviour design, and could be expected to correlate with robot persuasiveness, but that cannot be assessed through direct application of the equivalent HHI measure. It is also clear that robot credibility is not independent from the overall context of *why* the robot is utilising such behaviours, who designed them and for what purpose etc. More generally, the findings presented in this chapter demonstrate the difficulty in measuring perception of a robot using subjective measures; likely somewhat due to participants being conflicted in their answers (answering based on logic/rationale rather than emotional response) or differences in how the robot is framed/assessed (e.g. as an autonomous social agent presented with no background or detail on proposed use vs. an extension of the programmer and specifically within the context of its ultimate application). This highlights the importance of objective measures based on user behaviour, and the value of qualitative data collection for generating further insight into participant responses.

3.6.3 Designing Socially Persuasive Robots, *Ethically*

Design implications for leveraging persuasive strategies in SAR design are presented at the end of this subsection. Given that these recommendations might be considered somewhat at odds with ethical guidelines for robot design, such as the recent BS 8611, the way this particular standard was considered during this work and the resulting rationale/justification for potentially going against it is presented first. It is important to note however that the ethics concerning SAR use more generally are not a focus of this work. Rather, this is a practical consideration of and attempt to apply emerging ethical guidelines to real world robot design and development.

Study 1 specifically utilised dialogue that actively opposed the mitigation strategies for avoiding deception and anthropomorphisation presented in BS 8611 and demonstrated why this might be worth doing in the context of a real world, useful SAR application. Qualitative data collected during this study suggest that participants were generally well aware of the robot's nature, e.g. that it had simply been programmed to act in a certain way and did not feel any emotion. Similarly, responses to the questions concerning relationship development to/from the robot further suggest that participants were acutely aware that the robot was unable to form relationships, as the distribution of answers to these two questions were significantly different, and the correlation between them was not as strong as might be expected. This suggests that participants were able to decouple the two, and identify that whilst they may feel something toward the robot, that was not necessarily mutual. Answers to the questions on genuineness and deception further show that, specifically in the context of SAR applications, the demonstrated socially persuasive behaviours were considered acceptable. This was also echoed in responses to the equivalent questions implemented for studies 2 and 3.

Studies 2 and 3 were specifically designed to (i) demonstrate what socially persuasive behaviour more aligned with the mitigation strategies presented in BS 8611 might ‘look’ like and (ii) generate preliminary results as to what impact this might have on the robot’s effectiveness in terms of credibility, user preference/acceptance etc. The results from Study 2 show that on demonstrating expertise or authority, having the robot refer to a credible, human third party rather than itself may be most beneficial. Further, results from Study 1 suggested that having the robot refer to and present its *own* expertise had no benefit for persuasiveness. Having this reference to an appropriate human authority would also align well to the general ethical principle of ensuring users are aware of who is ultimately responsible for the robot’s behaviour, as well as being less anthropomorphic and hence deceptive. So for expertise, more ethical interaction design is likely to also be more effective. On demonstrating social behaviour however, the Study 3 results on robot credibility show that for the more emotive/interaction based socially persuasive behaviours (i.e. showing empathy, goodwill etc.) it may be more beneficial for the robot to present *itself* as the social agent behind those, rather than referencing others.

However, qualitative data collected from both Study 1 and Study 3 demonstrate that participant opinions on the appropriate style and amount of social interaction for this kind of application is very much a user personal preference. In Study 3, whilst the robot with the more anthropomorphic dialogue was most commonly identified as one participants would rather work with, this was not by such a large margin as the human versus robot expertise condition robots of Study 2. In fact, a notable number of participants specifically indicated they would prefer to work with the control robot which had basically *no* social interaction at all, because they felt it was more efficient and preferred the lack of pretence.

Similarly in Study 1, the overlap in themes emerging from participants likes and dislikes regarding the robot, with social interaction being something some participants liked but others disliked, shows that the same behaviour can be perceived completely differently by different users. This points again to the need for personalised approaches to socially persuasive behaviour, which is in line with the results regarding personalisation of therapy/behaviour presented in Chapter 2. Persuasion models from HHI might be able to initially inform such personalisation. For example, the ELM specifically identifies that the amount of rationale provided versus peripheral cue manipulation attempted by a persuader should be informed by an understanding of the receiver’s elaboration level. However, socially intelligent assessment and tracking of such a complicated construct for informing appropriate and effective persuasive strategies represents another example of the kind of human expert, intuitive/experience-based skills that need to be replicated by SARs in order for them to be effective.

In summary, the findings presented in this chapter suggest that (i) anthropomorphic and potentially deceptive behaviours based on persuasion strategies from HHI literature can improve the effectiveness of a SAR, (ii) personalisation of such behaviours will likely be key in maximising and maintaining such effectiveness (particularly over the long-term), (iii) such behaviours may

not *actually* deceive the user with regards to the robot's nature and (iv) the use of such behaviours *in the context of SARs* is acceptable to users. It must be noted that these findings are limited to the non-vulnerable, adult population from which participants for these studies were drawn. However, combined with the results from the study with therapists in Chapter 2, they provide initial justification that SARs might represent one of the '*well defined and socially-accepted purposes*' for actively utilising anthropomorphisation referred to in BS 8611. With these points in mind, specific strategies for designing socially persuasive SARs should include:

- Having the robot show an interest in the user *i.e. the robot should ask about the users' feelings and or wellbeing, e.g. with regards to the task*
- Having the robot suggest some sympathy/empathy *i.e. based on the above, the robot should respond with an appropriate acknowledgement, of matched emotional valence*
- Having the robot demonstrate some similarity to the participant *i.e. the robot should indicate it shares the users' preferences regarding the task/topic*
- Ensure that the expectations set by the above behaviours, if deployed in pre-task interaction, are then met by the successive interactions, particularly around task encouragement and the robot's perceived engagement in/understanding of the participant's task performance *i.e. after displaying an affective interest in the user the robot should continue to provide social encouragements and interactions during execution of the task*
- Doing all of the above in an anthropomorphic '*first person*' way *i.e. presenting itself as the social agent actively engaging in those social behaviours rather than referencing human third-parties*
- Having appropriate persons, already perceived as highly credible to users (e.g. therapists), introduce and/or endorse the robot
- Refer to that/those human(s) when demonstrating expertise/authority *i.e. when making a task-related request or providing task rationale*
- Adaptive tailoring and personalisation of persuasive strategy and related social interactions/behaviours to the user and the context

3.7 Conclusion

This chapter presents findings from three studies concerning the effectiveness and acceptability of socially persuasive behaviours, inspired by results from the previously presented study with therapists and an informed review of HHI literature on persuasion, in the context of SARs. Together, the studies consider both pure, objective '*effectiveness*' of these behaviours *i.e.* how they

can impact on user behaviour, but also the potential for ethical risk and what impact trying to minimise this might have on that effectiveness. Key findings across the combined results can be summarised as follows:

- Demonstrations of goodwill and similarity increased the persuasiveness of a SAR, as measured objectively by the amount of voluntary exercise participants elected to complete.
- Demonstrations of expertise did not have the same effect, but this may be because such demonstrations are only effective for robots if they refer to the *human* source of such expertise.
- Based on the above, the Elaboration Likelihood Model (ELM) of persuasion seems to be an appropriate model for interactions whereby a SAR is designed to encourage or guide particular user tasks/behaviours, offering inspiration for robot behaviour design.
- Robot *credibility* does seem to be a construct that exists in perceptions of a social robot, and that can be impacted by that robot's social behaviour. However, robot credibility is not independent from the credibility of the robot's programmer, associated expert humans and/or the context of application. Predominantly due to this, typical HHI based questionnaire measures for credibility are likely not appropriate for assessing robot credibility, and more generally the use of subjective questionnaire measures such as the Godspeed questionnaire should be used with caution in HRI studies.
- Building on the above, user acceptability of socially persuasive behaviours is explicitly linked with their purpose. Within the context of SARs, there seems to be a consensus that even if such behaviours *are* deceptive (many participants suggested they were not) then that deception is acceptable because these behaviours are particularly appropriate and/or effective for such use cases.
- Whilst in some cases it might be that designing more ethical social robot behaviours also results in them being more effective (e.g. with displays of expertise) it is likely that SAR behaviours will need to leverage anthropomorphism to be effective and will as such go somewhat against emerging ethical standards concerning the need to minimise this. However, the positive acceptability and lack of deceptiveness results noted above suggest that SARs for exercise encouragement might be one '*well-defined and socially acceptable application*' (BSI 2016) for which such use of anthropomorphisation is justified.

These findings were analysed to generate a set of design implications, for informing SAR and persuasive social HRI design more generally, that include a practical attempt to incorporate compliance with/awareness of the recent British Standard BS 8611 for the ethical design of robots and robotic devices. It is also hoped that these design implications and the discussion surrounding key results in this chapter demonstrate the benefits of (i) objective rather than

subjective measures and (ii) qualitative data collection in the context of understanding social HRI, thereby motivating future works to incorporate these into any experimental protocol.

Concerning the ultimate goal of this research, the design and automation of an autonomous SAR, the results in this chapter have the following implications:

Importance (and Complexity) of Personalising Persuasive Social Robot Behaviour

The different persuasion strategies identified for high and low elaboration by the ELM, some of the very mixed results from this work regarding participant preferences, perception and acceptability of behaviours, and results from the study with therapists in Chapter 2 all highlight the importance of personalisation, but also the complexity in doing that effectively. Neither the HHI literature reviewed nor the study with therapists resulted in explicit and tangible ‘rules’ for informing the personalisation of persuasive behaviour (beyond the high versus low elaboration strategies identified by the ELM) in a form that could be used to automate SAR behaviour. The study with therapists specifically suggested that this type of personalisation was only possible after ‘getting to know’ the service user. This points towards the potential of an expert-in-the-loop approach to automation that (i) specifically allows an expert to monitor a SAR interaction, over time (to allow for personalisation effects) and (ii) allows the expert to provide feedback in real time based on what ‘feels’ right (rather than having to explicitly provide some sort of heuristic or reasoning). This therefore motivates the expert-in-the-loop machine learning approach, applied to automation of a SAR, presented in Chapter 4: *Creating an Autonomous, Socially Assistive Robot with Interactive Machine Learning*.

Interactions Between the Robot, its Application and Those Behind its Deployment

Particularly regarding the acceptability of socially persuasive robot behaviour, the results in this and the previous chapter demonstrate the inability to consider a robot *in isolation*; i.e. outside of contextual factors surrounding the intent behind its design, proposed application and environment into which it would be deployed. This is further evidence of the need to take a mutual shaping approach to robot design and development, as employed throughout this work and reflected upon in Chapter 5 *Mutual Shaping in Design and Deployment of Socially Assistive Robots*. More specifically, the potential for *inherited credibility*, or acceptability and trust of socially persuasive behaviours being linked with them being signed off by an expert, further motivates the use of expert-in-the-loop methods like that employed in Chapter 4.

CREATING AN AUTONOMOUS, SOCIALLY ASSISTIVE ROBOT WITH INTERACTIVE MACHINE LEARNING

Results from the studies presented in Chapters 2 and 3 offer a number of design guidelines for creating an effective socially assistive robot (SAR). The next step then becomes a consideration of how to automate socially assistive robot behaviour, utilising these guidelines, in order to create a real world useful system. The study with therapists in particular identified that whilst experts can identify things they take into consideration (i.e their *inputs*) and the different behaviours they may exhibit in different scenarios (i.e. their *outputs*), it is difficult for them to explain the link between them. There were many references to the importance of *getting to know the client* and acting based on *intuition and experience*. This chapter presents the design, implementation and evaluation of an interactive machine learning (IML) system designed with this in mind. The IML setup allows a domain expert to contribute to automation of robot behaviours directly, by observing the robot in action and providing training data to the learning system in real-time. This work and a subset of the results presented in this chapter will be presented at Robotics: Science and Systems 2020.

4.1 Introduction

The results from Chapter 3 demonstrated that having a robot demonstrate human-like social persuasive strategies was effective in motivating engagement with a task. However, all work in that chapter utilised a wizard-of-oz setup and hence did not consider how such behaviours might be made autonomous. Many previous works on automating social robot behaviour have attempted to emulate lifelike behaviour by employing models based on human or animal psychology (e.g. Lemaignan et al. (2017), Arkin et al. (2001)), or through observing and then attempting

to replicate human-human interaction (Sussenbach et al. 2014). Such methods are relatively inaccessible to non-roboticists, which limits the potential for direct input from domain experts (teachers, therapists etc.) who are skilled in the use of social interaction in complex assistance scenarios. They also offer little opportunity to consider the dynamic interaction between robots and their context of use; noted throughout this work as crucial for the successful real world deployment of SARs.

An alternative approach is to have a human *teach* the robot how to behave. This is typically achieved using Learning from Demonstration methods in which a human controls the robot to demonstrate the desired behaviour, resulting in a training dataset to which machine learning can then be applied offline (e.g. Knox et al. (2014), Sequeira et al. (2016), Clark-Turner & Begum (2018)). State of the art work has gone past this to utilise interactive machine learning (IML) which allows the learning process to occur in real-time, such that the robot can be used and trained simultaneously. Such an approach was recently demonstrated as a feasible method for generating autonomous, socially assistive robot behaviour via the framework of Supervised Progressively Autonomous Robot Control (SPARC) (Senft et al. 2019). The following subsection gives more detail on how SPARC works and why it was identified as a promising methodology to employ in this work.

4.1.1 Interactive Machine Learning for Automating SAR Behaviour

Figure 4.1 provides a simplified overview of the interaction flow underpinning the IML element of SPARC. During training/supervised use of the robot, the robot interacts with the participant directly, but under the close supervision of the expert. Initially, the robot does not have any action policy, and the expert effectively teleoperates the robot by providing action *exemplars* that are directly executed by the robot, and simultaneously added to the training dataset. After each example, the robot incrementally trains its own model, in order to progressively learn its own action policy. Early on in the process, the robot starts to generate action suggestions, that are sent to the expert for validation. If accepted, these suggestions are positively rewarded; if rejected, they are negatively rewarded. Combined with the expert-initiated exemplars, this helps the robot machine algorithm to quickly converge toward an appropriate action policy. Once the expert is confident that the robot's suggestions are 'good enough to be trusted', they can 'switch' to the autonomous mode, in which the robot's suggestions are automatically accepted, without any human intervention, resulting in a fully autonomous behaviour.

Overall, the SPARC approach was specifically identified as being appropriate for this work for two key reasons. Firstly, it provides a mechanism for learning from intuitive expert behaviour and tacit expertise that are difficult to verbalise. Secondly, as a methodology, it aligns with the expert-informed and mutual shaping approach taken throughout this thesis. In further detail, SPARC is an appropriate and advantageous method for automating SAR behaviour because:

- (i) it requires only the identification of the robot's input and output space (which can be

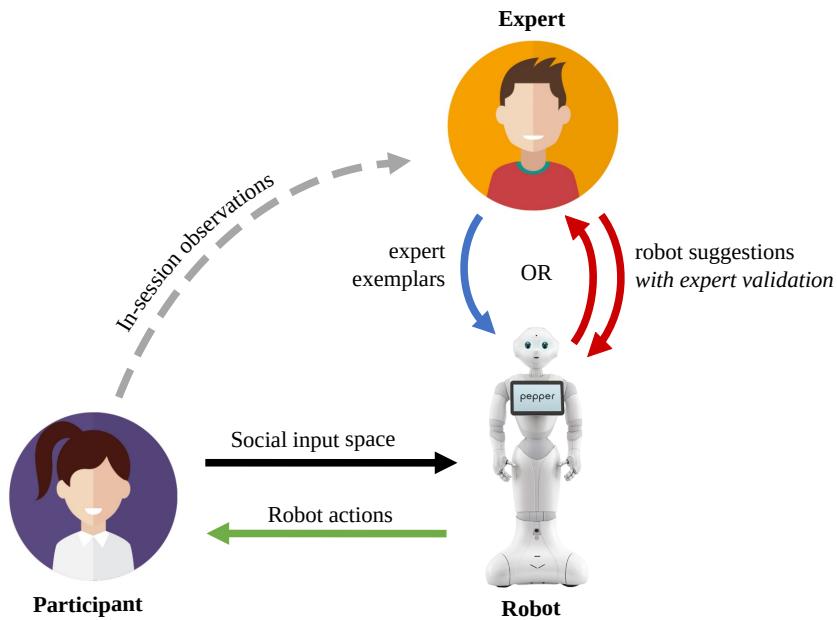


Figure 4.1: Expert-in-the-loop machine learning: during the robot-participant interaction, the expert teacher can either *initiate* suitable robot actions (expert exemplars), or the robot itself can *suggest* actions, that are then *evaluated* by the expert. Every time an action is initiated or evaluated, a tuple {*input space; action; reward*} is added to the training set of the robot's learning algorithm. Over time, the robot suggests more and more appropriate actions, and becomes progressively autonomous.

co-designed with a domain expert) without any rules between them

- (ii) it allows a domain expert to contribute to design *and* automation of the robot
- (iii) building on the above, it offers a specific path to personalisation of robot behaviour
- (iv) deployed in the ‘real world’ it can be responsive to mutual shaping effects (reflected in the expert’s control/supervision of the system)
- (v) having an expert-in-the-loop during both training and autonomous operation is desirable from both a safety and ethical perspective (concerning e.g. the verification/validation of resultant autonomous behaviours and prevention of any inappropriate behaviour)

This chapter describes development and application of the SPARC methodology to an exemplar use case representative of the long-term, monotonous exercise engagement scenarios considered in Chapters 2 and 3. This is described in detail below. Results presented in this chapter demonstrate that the approach was fundamentally successful in generating appropriate autonomous behaviour. Chapter 5 further details how this methodology can be considered as a participatory design technique. Additional results and observations presented in that chapter further demonstrate that this approach supports pursuit of a mutual shaping approach to robot design, development and deployment.

4.1.2 Couch to 5km: a Real World SAR Application

Given the difficulty in identifying a suitable therapist-patient group into which a developmental SAR could be (safely) deployed and tested, it was decided to instead target the similar but lower risk use case of a long-term exercise programme designed for individuals wanting to become more active. Specifically, the UK National Health Service (NHS) Couch to 5km (C25K) programme was identified as an appropriate proxy use case because it is:

- a long-term exercise programme requiring a relatively high level of commitment from participants
- a relatively boring exercise programme consisting of running and walking only
- similar to a rehabilitation programme in that it is of fixed duration with a clear aim of achieving a specific performance goal in that time

The programme consists of undertaking 3x weekly exercise sessions for 9 weeks, with sessions building up from a combination of short runs and walks to a full 30 minutes running, and is currently delivered via podcast¹. The simplicity (and relatively boring nature) of the task places great importance on the ability of the ‘C25K coach’ robot to provide engaging and appropriate social supporting behaviour. In tackling the C25K programme, this work also addresses two key interaction features not yet considered in the previous application of SPARC: long-term interaction and personalisation of robot behaviour, such that the resultant system should be able to identify *what* actions to do, *when* and for *who*.

4.2 Technical Approach

The exact role of robot, as well as its action and input spaces, were co-designed with a domain expert (a fitness instructor from the University of the West of England Centre for Sport) over 6 co-design sessions between the author, fitness instructor and occasionally an additional member of the supervisory team. These sessions were conducted over a period of 5 weeks and represent a total of 12.5 hours direct co-design work as described in Table 4.1. More detail on these sessions, and how they fit into the larger ‘*IML as participatory design*’ process posited by this work, is given in Chapter 5.

4.2.1 Assumptions and Limitations

Clearly the interactions between an exerciser, their exercise programme and their fitness instructor are hugely complex. Studies looking at factors affecting adherence to a programme have identified a huge range of factors that might influence both immediate and long-term engagement

¹<https://www.nhs.uk/live-well/exercise/get-running-with-couch-to-5k/>

Session (length)	Design Activities
1 (2 hours)	Interaction scenario was presented to fitness instructor for initial/unbiased recommendations, researcher then shared pre-prepared suggestions based on previous work to brainstorm initial key actions/inputs
2 (2.5 hours)	Visited gym that will be used for the study
3 (2 hours)	Fitness instructor conducted mock Couch to 5km session with a supervisor observed and filmed by the author Physical prototyping of the teaching interface tablet Took draft tablet teaching interface to the gym, fitness instructor put the author through mock session choosing actions via the tablet interface and verbalising what he was doing/why (action choices were stored via the tablet interface and the instructor's use of the tablet was also video recorded)
4 (2 hours)	Author and fitness instructor went through resulting video footage to further discuss what participant information may have been informing his action choice Went through the action space to discuss specific utterances/examples for each action Produced heuristics for a rule-based, autonomous system (<i>discussed in Chapter 5</i>)
5 (1 hour)	Dictionary of specific utterances worked on predominantly by the fitness instructor alone
6 (3 hours)	Tested the experimental set-up with the fitness instructor trying out both the teaching and participant role Made final practical study decisions e.g. fitness instructor placement during the running sessions etc.

Table 4.1: Key participatory/co-design activities undertaken between the research team and domain expert to develop the IML robot set-up including the naive robot action space, input space and teaching interface.

with exercise, from a lack of belief about outcomes, depression and lack of interest to the weather on a given day (O’Shea et al. 2007, Forkan et al. 2006). Similarly, the range of ways in which a fitness instructor or other social presence can support and positively impact on that engagement cannot simply be reduced to the idea of one role/persona with a discrete list of supporting actions. Whilst a lot of care was taken to closely co-design the proposed system with the fitness instructor, a number of simplifications and assumptions were made in framing the problem, and a number of limitations result from the approaches employed. These reflect the overall scope, SAR framing and research philosophy underpinning all of this work, as set out in Chapter 1.

The Role of the C25K Robot Coach

As described previously, the experimental interaction scenario, and the role of the robot within that, was carefully co-designed with the fitness instructor. It was further informed by (i) results from the studies presented in Chapters 2 and 3 and (ii) the overall mutual shaping/responsible innovation approach taken throughout the work. This resulted in the following key decisions and assumptions regarding the role of the robot in the context of delivering the C25K programme:

1. Ultimately the purpose of a C25K robot coach would be to *assist* rather than *replace* the fitness instructor.

As made clear in Chapter 1, the aim of this work has never been to replace humans in the context of socially assistive scenarios. Rather, the focus has been on understanding how SARs might assist these vital workers in a meaningful way, such that the robot is a tool to support human, expert led interventions. For the C25K robot coach specifically this would mean that, during the programme defined exercise sessions, the robot would ultimately be able to direct and encourage the pre-defined exercises with very little input required from the supervising instructor. In turn, this assumes that in practice the instructor will

actually be open to ‘handing over’ some responsibility to the robot such that his workload would reduce as expected.

2. Building on the above, the role of the fitness instructor (and his visibility) is important in the context of delivering the programme.

Whilst HRI studies typically attempt to minimise the presence and therefore potential impact of human (e.g. researcher) presence, in this study it is acknowledged and actually desired from the beginning that the role of the fitness instructor is somewhat explicit and clear to the participants. Again, this results primarily from the overall aim of this work regarding development of robots that would only ever form *part* of a human-led intervention. Results will need to be carefully considered to understand the impact of the robot and the instructor as independent social agents, but those results will be valid regarding how robots might realistically be deployed, alongside humans, in the real world.

3. Whilst also providing social support, the robot will be explicitly providing task instructions and therefore takes a somewhat authoritative role with regards to the programme.

Like the studies in Chapters 2 and 3, the C25K scenario requires the robot to guide users through a series of prescribed tasks that they may have little motivation to do. As the robot provides the task instructions and related encouragements, it fundamentally attempts to be somewhat authoritative to the participant. It is important again to recognise that this is just one role/dynamic a SAR might take with regards to effecting user behaviour change through social influence, and will not be appropriate for all use cases/all users in other socially assistive scenarios.

Lack of Technical Running Assessment / Feedback

It was decided early on that the robot coach would not attempt to assess and provide feedback on participants’ specific running technique. This would require complex capture and kinematic modelling of participants’ movements, requiring additional, non-trivial technical development. In addition, giving such feedback was considered to be high risk, in that incorrect feedback could lead to participant injury. As such, it was decided that correcting technique or providing other specific advice like this would be left to the fitness instructor, only to be done *outside* of exercise sessions unless a participant was actively at risk when they were exercising with the robot.

The Consistency of Expert Demonstrations

The expert-in-the-loop IML approach inherently assumes that the expert teacher will be consistent in their use of the system and hence their generation of training data. Some variation is to be expected, e.g. different types of actions might become more appropriate as the programme progresses. This is ‘*good*’ variation that it is *desirable* to capture with the IML system. However, successful learning of such variation requires at least some of the factors informing these changes

(e.g. progression through programme) to be included in the system input space, which requires successful identification of them ahead of system deployment.

In addition, outside of these factors, the consistency of the expert is not guaranteed. Just like participants, the expert's interactions with the system can be affected by short term factors such as mood and fatigue, and over the longer term will no doubt change over time based on familiarity with the system, increased understanding of its potential etc. Further, it seems likely that as the IML system starts generating suggested actions, these might influence the expert to e.g. accept a suggestion even if it wasn't an action *they would* have executed had they not been prompted. The underlying assumption is therefore that, over time and averaged out across numerous training examples, the expert will ultimately be guided (albeit subconsciously) by a fairly consistent action policy that the IML system will learn to replicate, and that this will not be unduly influenced by suggestions coming from the IML system itself.

Training and Evaluating IML within the Same Study

In this work, a single study (representing end-to-end delivery of the fixed term Couch to 5km programme) is used both to train the IML system and test its resultant autonomous behaviour. This means that the autonomous behaviour of the system is only tested in sessions that are late within the study and Couch to 5km programme. Engagement in the programme at this point, as well as evaluation of the robot, will likely be shaped at least in part by the long-term interactions participants have had with the system/instructor. Similarly, quantitative evaluation of this autonomous behaviour will require it to be compared to similar supervised sessions, also late in the programme. As such, this study set up doesn't allow for any testing of how well the system would autonomously perform in earlier sessions in the programme (which have more walks and shorter bursts of running) nor if/how well the autonomous system could generalise to new participants.

4.2.2 Modelling the Robot Action Space

Conceptually, it was identified that the robot coach would provide two key action types:

1. Task Actions

Those which provide direct task-specific instructions to the user, i.e. when to transition between walking and running according to the C25K programme.

2. Social Support Actions

Those designed to facilitate and encourage user engagement with the task (essentially socially persuasive behaviours leveraging the social influence of the robot) e.g. celebrating participant effort, demonstrations of sympathy etc. Note these could also include low level/non-verbal behaviours, e.g. a change in proxemics.

CHAPTER 4. CREATING AN AUTONOMOUS, SOCIALLY ASSISTIVE ROBOT WITH
INTERACTIVE MACHINE LEARNING

Social Supporting Actions									Task Actions	Low Level	
	Time	Social	Performance	Reward	Check User	Animation	Get Closer	Back Off	Run	Walk	Eye Colour
P	Time	Humour	Maintain	Praise	-	Animation	-	-	Run	Walk	Green
C	Time	Challenge	Speed Up	-	-	-	-	-	Run	Walk	Yellow
S	Time	Challenge Sympathise	Speed Down	Praise	Check PRE	-	-	-	Run	Walk	Blue
N	-	-	-	-	-	-	Get Closer	Back Off	Run	Walk	White

Table 4.2: Full listing of Task and Social Supporting Actions for the C25K robot coach, all of which can be described by a *{action-type, style-modifier}* pairing, plus application of style modifiers to eye colour as an example of using Style to modify low level behaviour independent of specific actions. ‘Check PRE’ = Check Perceived Rate of Exertion. Style modifiers: P = Positive; C = Challenging; S = Sympathetic; N = Neutral.

Action Type	Description
Time	Encouragingly referring to the amount of time remaining on the run/walk
Social	A demonstration of ‘social support’/interaction such as giving encouragement or telling a joke
Performance	Giving task-specific feedback to speed up, down or stay the same
Reward	Praising the user for their effort/performance
Check PRE	Checking the user’s Perceived Rate of Exertion (PRE); asking the user how they are feeling (with user response submitted via icons displayed on the chest-mounted touch screen tablet)
Animation	Performing one of Pepper’s stock, positively valenced, non-verbal animations (utilising sound, movement and eye colour)
Get Closer	Pepper ‘leans’ forward (chest tilts forward from the waist)
Back Off	Pepper ‘leans’ backward (chest tilts backward from the waist)
Run	Introducing the length of the next run and counting down into the transition from walking to running
Walk	Introducing the length of the next walk and counting down into the transition from running to walking
Eye Colour	Changing of Pepper’s eye colour

Table 4.3: A brief description of each of the robot’s action-types.

These actions can be additionally described and/or shaped by an overall *mood* or *style*. As such, both Task and Social Support actions can be described by an *{action-type, style-modifier}* pairing, as shown in Table 4.2. For example, a *{Run, Sympathetic}* action would have the robot say “*Ok now I know you can do this, next is a run for 5 minutes*” whereas a *{Run, Challenge}* action would have the robot say “*Right I want to see you push hard on this next run for 5 minutes*” and a *{Social, Positive}* pairing results in a *Humour* action which might have the robot say “*You can call me Terminator because I’m going to make you run for your life!*”. Further, it was noted that style-modifiers could also inform lower level, non-action specific robot behaviour such as proxemics and non-verbal communication cues. As a demonstrator for this work, styles were used to set robot eye colour as shown in Table 4.2.

A brief description of each robot action-type is given in Table 4.3 and example dialogue for each speech based *{action, style}* pairing is given in Table 4.4. Note that a dictionary of utterances was created for each speech based action, with these utterances being cycled through each time the action was used.

Action-Type	Style	(Action)	Example
Time	Positive	Time	<i>You are flying through this run!!</i>
	Challenging	Time	<i>Don't you dare give up on me now, just a little longer!</i>
	Sympathetic	Time	<i>Almost there, I know you've got this!</i>
Social	Positive	Humour	<i>Run! Run like you're escaping the robotic revolution</i>
	Challenging	Challenge	<i>Come on [name] show me what you're made of!</i>
	Sympathetic	Sympathise	<i>It doesn't matter how you finish, even a bad run has its benefits</i>
	Sympathetic	Challenge	<i>I'm afraid it doesn't get easier, you only get tougher</i>
Performance	Positive	Maintain	<i>Great work, now keep that pace!</i>
	Challenging	Speed Up	<i>Come on you can work harder than this! Could you turn the speed up?</i>
	Sympathetic	Speed Down	<i>Woah there, let's not burn out too fast...</i>
Reward	Positive	Praise	<i>I'm impressed, you're doing great!</i>
	Sympathetic	Praise	<i>I know this is hard but you've got it</i>
Check User	Sympathetic	Check User	<i>How are you holding up [name] ?</i>
Run	Positive	Run	<i>Keep up this effort, now it's time to run for [time]</i>
	Challenging	Run	<i>Come on now let's push hard on this next run for [time]</i>
	Sympathetic	Run	<i>Come on you've got this, next up is a run for [time]</i>
	Neutral	Run	<i>Ok, let's switch to running for [time]</i>
Walk	Positive	Walk	<i>Great job! Now let's switch back to walking for [time]</i>
	Challenging	Walk	<i>Keep it strong now with a fast walk for [time]</i>
	Sympathetic	Walk	<i>Ok, I know that was hard, let's switch back to walking for [time]</i>
	Neutral	Walk	<i>Time to think about slowing it down now. Bring it down to a walk for [time]</i>

Table 4.4: Example dialogue for each speech based *action-type*, *style* action pairing implemented for the C2K robot coach. Note that a dictionary of utterances was created for each speech based action, with these utterances being cycled through each time the action was used.

4.2.3 Interaction Features and Input Space

Four key interaction features were identified in considering what input space might be required to inform a socially intelligent action policy covering both Task and Social Supporting Actions:

1. Task State

Describes non-performance related task information e.g. timing information, task the user should be undertaking (walking or running).

2. Task Performance.

Describes if/how well the user is performing the prescribed task.

3. Task Engagement

Captures to what extent the user is engaged with the prescribed task, recognising effort as being distinct from performance. Also note that this can include instantaneous as well as longer-term measures (e.g. speed during sessions versus overall attitude towards task outside of sessions).

4. User Personality

e.g. Big Five personality traits (Gosling et al. 2003) and other measures of attitude/motivation.

Full consideration of these categories requires an input space that combines *static* data, which does not change over the course of interaction, and *dynamic* data, which updates during the

CHAPTER 4. CREATING AN AUTONOMOUS, SOCIALLY ASSISTIVE ROBOT WITH INTERACTIVE MACHINE LEARNING

Type	Feature	Values	Description
Task State	Task Action Type	0, 0.5, 1	Whether participant is in warm-up, walk or run
	Session Progress	0-1	Time spent in session/session duration
	Programme Progress	0-1	Time spent on programme/programme duration
	Programme Action Progress	0-1	Time spent on current walk or run action/action duration
	Programme Action Duration	0, 0.5, 1	Current walk/run action length as \leq 3 mins, \geq 20 mins or other
	Time Since Last Action	0-1	Time since last action/60; capped at 1
Task Performance	Relative Speed: Average	0-1	Current speed/(2 x average speed)
	Relative Speed: Best	0-1	Current speed/(2 x personal best speed)
Task Engagement (Dynamic)	Heart Rate	0-1	Heart rate/2x resting heart rate capped at 1
	Motivation/Effort	0, 0.5, 1	Self-reported measure in warmup/on check PRE action
	Facial Expression: Lip Pull*	0-1	Normalised action unit returned by OpenFace
	Facial Expression: Mouth Open*	0-1	Normalised action unit returned by OpenFace
Task Engagement (Static)	Elaboration level (self)	0-1	Normalised sum of 3 Likert questions (derived from [anon. ref])
	Elaboration level (expert)	0-1	as above but rated by fitness instructor
	Activity Level	0-1	Likert question response
User Personality	Extroversion	0-1	Big Five measure normalised with respect to max score
	Agreeableness	0-1	Big Five measure normalised with respect to max score
	Conscientiousness	0-1	Big Five measure normalised with respect to max score
	Emotional Stability	0-1	Big Five measure normalised with respect to max score
	Openness to Experience	0-1	Big Five measure normalised with respect to max score

Table 4.5: Input space of the 20 state features implemented for the dual learning system (both style and action class learners utilised the same input space). The facial features marked * were later removed due to unreliability during testing.

interaction, demonstrated by the features listed in Table 4.5. Of particular interest are the *Task Engagement* measures, which include both static and dynamic measures. Specifically, these aim to capture an overall motivation with regards to the task and longer-term interaction scenario, as well as more instantaneous, in-session engagement/effort.

4.2.4 Dual Learner System

The composition of actions as *{action-type, style-modifier}* lends itself to a dual learning system, with the overall *learner* actually containing a ‘*style learner*’ and ‘*action learner*’; with each of these learners wrapping a classification algorithm suitable for use in IML. The output of these can be combined to generate actions, and the output of the style learner specifically can additionally be applied directly to low level behaviours as per Table 4.2. The overall information flow from the fitness instructor, into the dual learning system and then to the robot behavioural outputs is depicted in Figure 4.2.

4.2.5 IML System Architecture

A simplified schematic of the system control architecture is shown in Figure 4.3. All nodes communicate through the Robot Operating System (ROS) (Quigley et al. 2009) with a number of custom, study-specific ROS message types being implemented to describe e.g. the different types of actions. Communication with/control of the Pepper robot is done using NAOqi, Softbank’s multi-platform operating system².

²doc.aldebaran.com/2-5/index_dev_guide.html

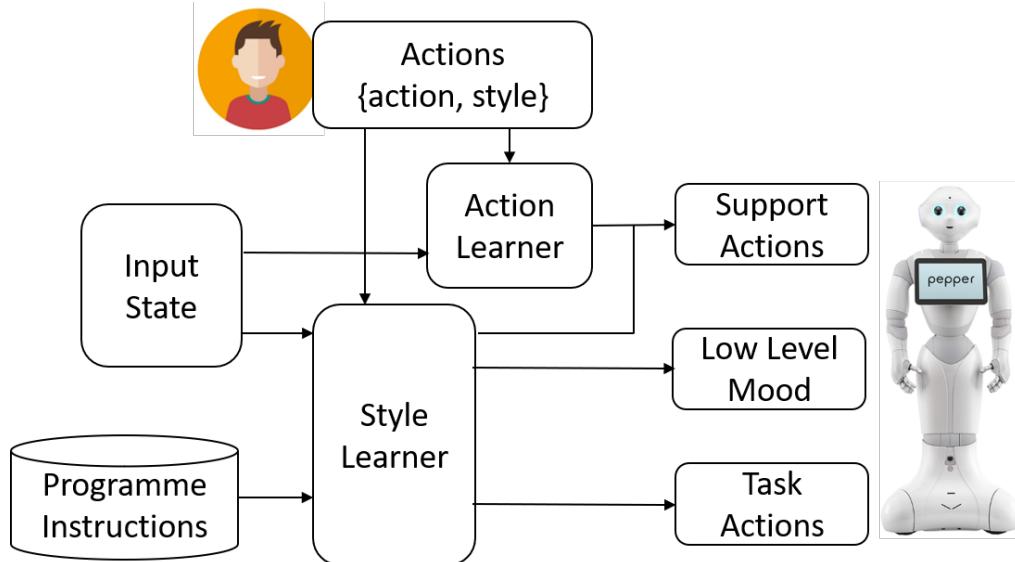


Figure 4.2: Overview of how expert initiated/evaluated actions are utilised in generating training data for each learner, and in turn how learner output is then utilised in generating robot behavioural output. Outputs from the Action and Style Learners are combined to create *{action-type, style-modifier}* pairings to result in actions, whereas output from the Style Learner can also be applied directly to (i) task actions and (ii) low level, non-verbal robot behaviours (i.e. eye colour for the C25K coach) to give an impression of the robot having an overall underlying ‘mood’.

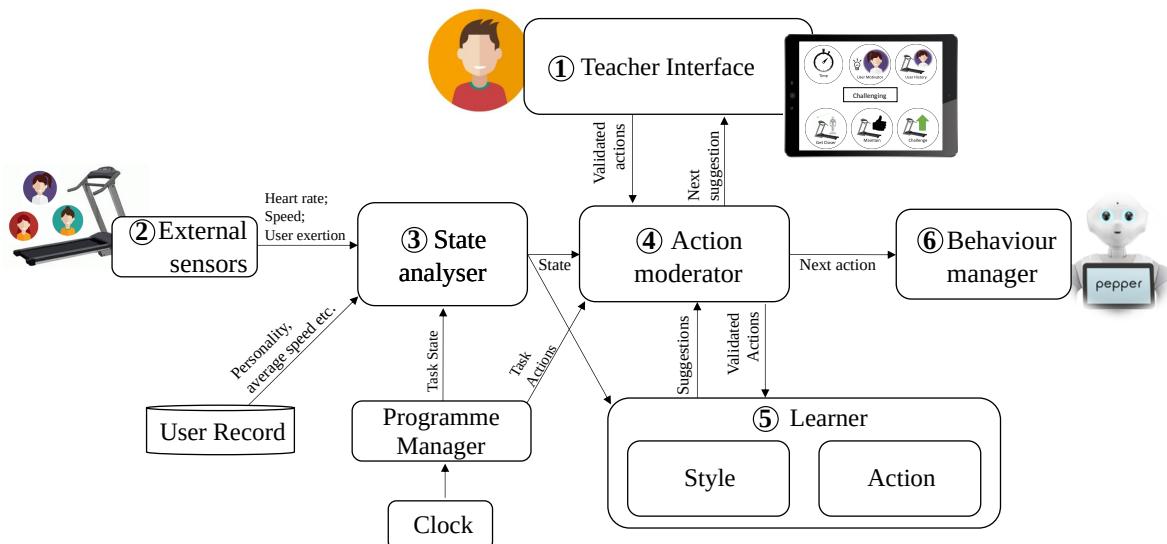


Figure 4.3: A representation of the system architecture used for learning from the fitness instructor and ultimately generating autonomous robot behaviour in guiding users through the C25K programme.

CHAPTER 4. CREATING AN AUTONOMOUS, SOCIALLY ASSISTIVE ROBOT WITH INTERACTIVE MACHINE LEARNING

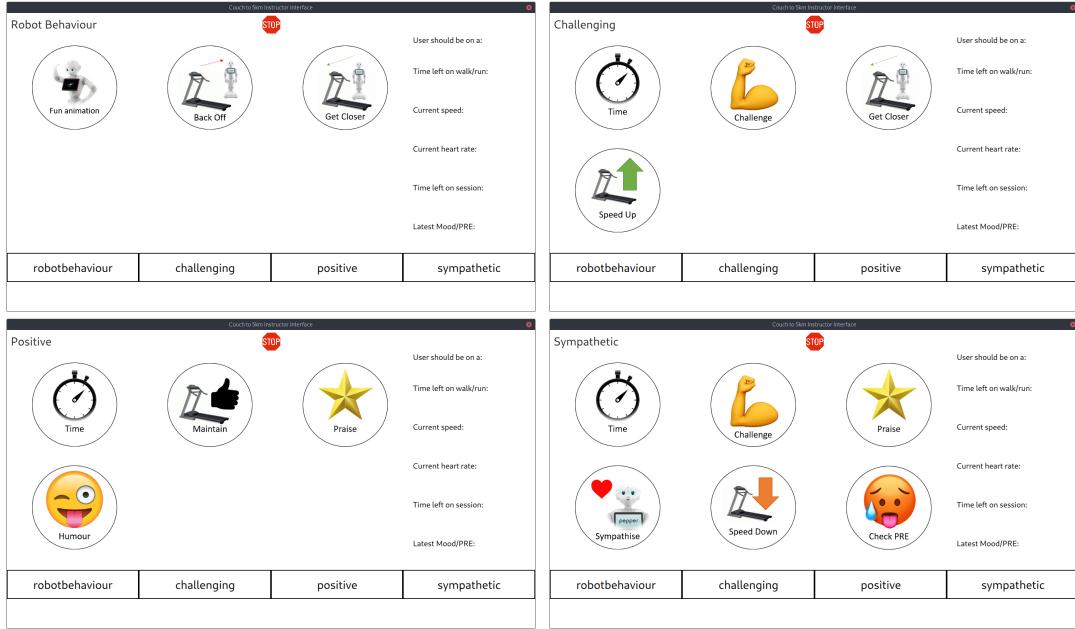


Figure 4.4: Pages of the teaching interface with Social Support actions (to be triggered as action exemplars, when necessary, by the instructor) organised by *style*.

4.2.5.1 Teacher Interface

The Teacher Interface (1) is used by the fitness instructor to i) initiate robot actions directly (providing *expert exemplars* as per Figure 4.1) and ii) respond to learner suggestions (including the suggested styling of Task Actions). Actions allowed to time-out at the interface, receiving no response from the fitness instructor within the given timeframe, are considered *passively accepted* and allowed to auto-execute. The teaching interface is coded in QML and runs on a touch-screen tablet held by the teacher during exercise sessions (shown in Figure 4.8). This interface was also co-designed with the instructor for maximum usability and ease of use during training/system supervision. Figure 4.4 shows the basic design of the tablet interface with regards to Social Support actions that the instructor could trigger as desired. Figure 4.5 then shows the tablet interface in use, presenting the instructor with (i) a Learner suggested Social Support action and (ii) Learner suggested styling of a Task Action.

4.2.5.2 External Sensors

Dynamic task data requiring the use of external sensors were addressed as follows:

Heart rate: captured via a polar H10 Bluetooth heart rate sensor³ worn on the user's chest. Raw output (in beats per minute) was displayed on the teacher interface, but for learning purposes was normalised with respect to user resting heart rate (see Table 4.5)

³polar.com/uk-en/products/accessories/polar_h10_heart_rate_sensor

4.2. TECHNICAL APPROACH

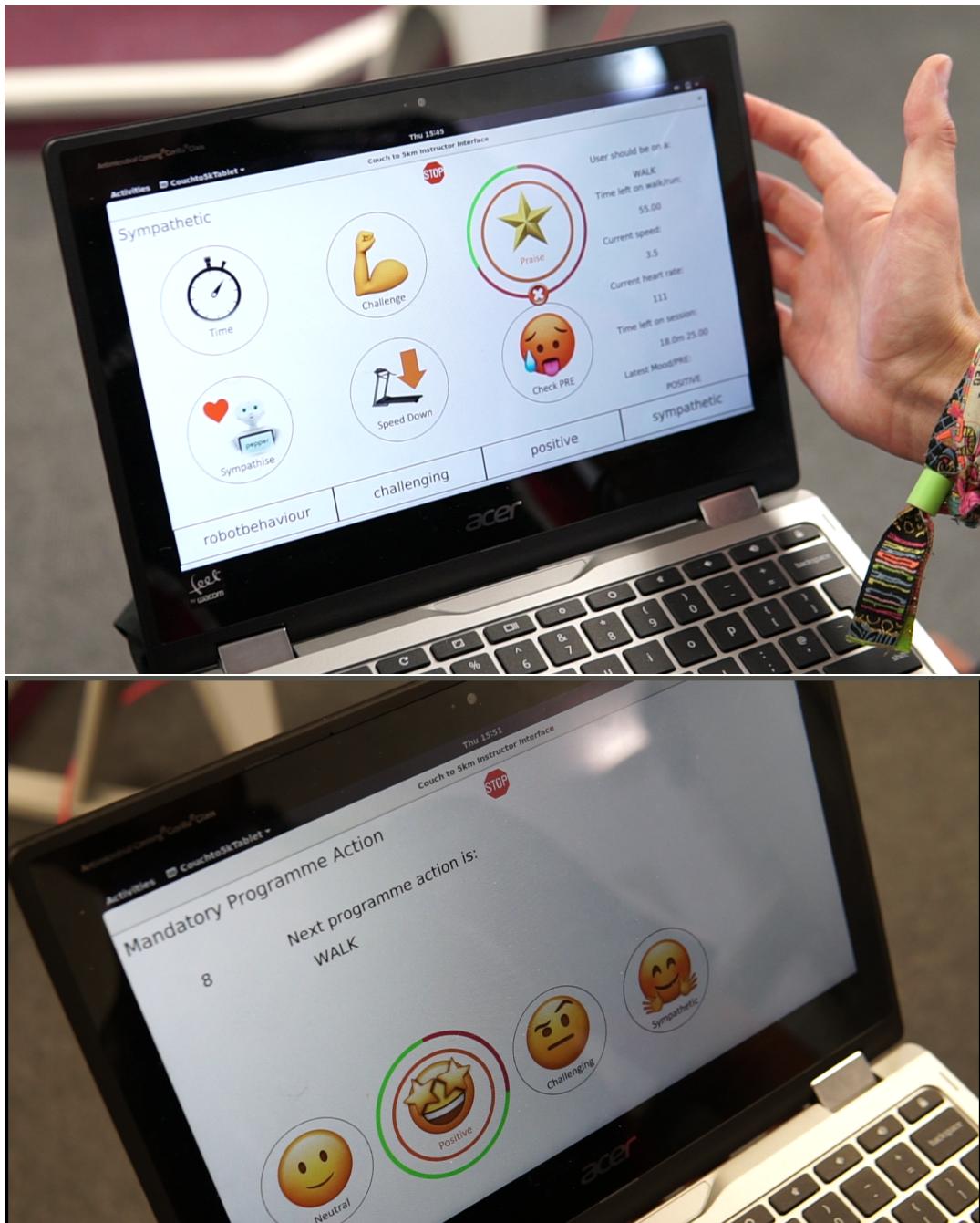


Figure 4.5: The C25K teacher interface in use, presenting the instructor with (i) a Learner suggested Social Support action and (ii) Learner suggested styling of a Task Action.

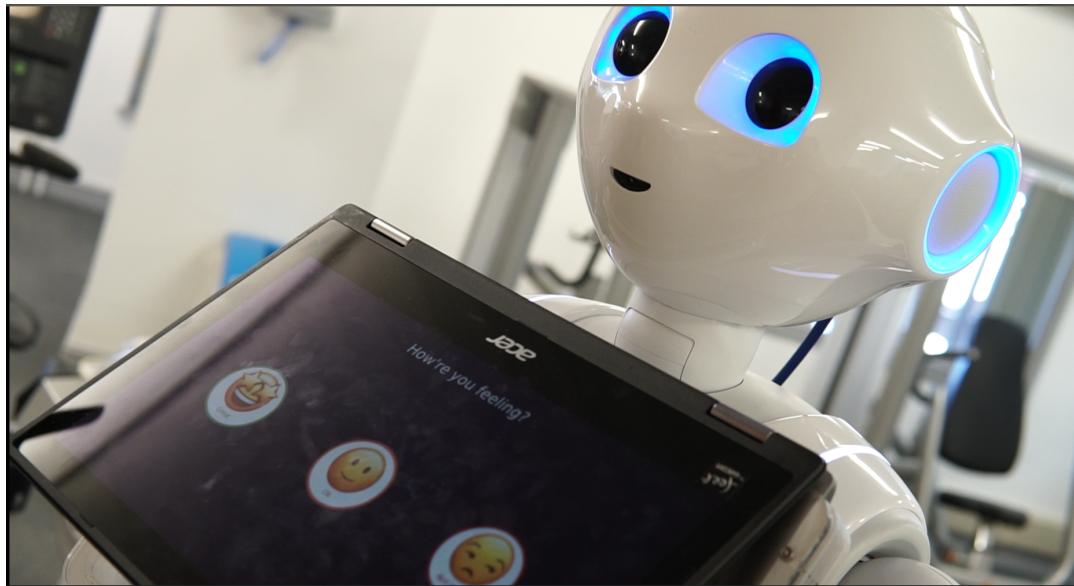


Figure 4.6: The C25K Robot-mounted tablet showing the ‘Check PRE’ action, with the robot’s speech in subtitles and icon-based buttons for the user to respond to its question.

Treadmill speed: read and automatically digitised from treadmill display using a treadmill mounted camera. Raw output (in miles per hour) was displayed on the teacher interface, but for learning purposes was normalised with respect to users’ average and personal best speeds, which were iteratively updated (at the end of every session) during robot deployment.

User Perceived Rate of Exertion (PRE): users were asked e.g. ‘How are you feeling’ by the robot, and asked to respond via the robot-mounted tablet (described below).

Facial expression: real-time extraction of ‘lip pull’ and ‘mouth open’ facial action units using OpenFace (Baltrusaitis et al. 2018) via a treadmill mounted camera. The robot’s camera was not utilised due to the robot’s positioning (the robot was not directly in participants’ line of sight - see Figures 4.7 and 4.8).

An external tablet was used in place of the robot’s tablet to allow for direct connection to the ROS architecture. A simple user interface was coded in QML, through which users could trigger start of sessions by selecting their user icon, respond to the robot’s Check PRE action (responding to e.g. ‘How are you feeling’ by selection an appropriate *good, ok, not great* response icon - see Figure 4.6) and view subtitles of the robot’s speech (published via ROS messages by the Behaviour Manager (5) at the time of action execution).

Additionally, there are two key databases within the architecture. Firstly, the User Record contains both static (pre-hoc collected data such as personality scores, activity level) and dynamic (e.g. time spent on programme, average and personal best speeds) user data. Secondly, the Programme Database (within the Programme Manager) identifies the timing of Task Actions for each exercise session according to the C25K programme.

4.2.5.3 State Analyser

The State Analyser (3) collects data from all input sources: external sensors, programme manager, internal clock etc. in order to produce the 20-dimensional input state used to describe the interaction (see Table 4.5). The state analyser receives data from sources at whatever rate those sources publish, but is set to publish specifically at 2Hz. This was selected as a reasonable base frequency that would allow for high responsiveness to changes in state without overwhelming the system. This also therefore represents the rate at which the Learner (5) is passed new input states and asked to make suggestions.

4.2.5.4 Action Moderator

The Action Moderator (4) facilitates the passing of actions between the fitness instructor, learner and robot. This includes:

- receiving Task Actions from the programme manager and applying the latest learner-suggested style before sending to the teacher interface for validation
- publishing appropriate *style* and *action* training examples to the Learner (5) based on actions initiated and/or validated at the Teacher Interface (1)
- relaying Learner (5) suggestions to the Teacher Interface (1) managing a priority queue based on action type and limiting the rate of action suggestion
- relaying fitness instructor initiated/accepted learner-suggested actions to the Behaviour Manager (6) managing a priority queue based on action type

In undertaking the above, the action moderator also accounts for timing requirements around (i) programme actions needing to be sent to the teacher tablet in advance of their programme-specified execution time and (ii) the time taken to actually validate and execute actions. Learner-suggested Social Support Actions suggestions made when the tablet was already displaying an action, or when the system was already executing an action, were disregarded. Task Actions were always given maximum priority and were never allowed to be disregarded by the queuing system as their timing was crucial to delivering the C25K programme.

‘Style Update’ Actions (for changing robot eye colour change) were generated by the Action Moderator when the Learner suggested style was different to the current style. These actions were automatically verified without fitness instructor approval (to reduce instructor workload), and had the lowest priority within the queuing system, so would be disregarded in favour of any Task or Social Supporting Action received by the Action Moderator ahead of their execution. As a reminder, the use of eye colour to denote current robot ‘mood’/‘style’ was selected as a simple proof of concept for using the Style Learner to inform low-level robot behaviour independent of Task and Social Support Actions.

4.2.5.5 Learner

The Learner (5) is responsible for accepting training examples and generating suggested styles and Social Support Actions. It wraps generalised machine learning algorithm instances (one for *style* and one for *action-type*), interpreting the input/output between them and the wider system. This makes it easy to switch between different classification algorithms. The Learner is event driven primarily by the receipt of training examples and input states, demonstrated by the example code excerpt presented in Algorithms 1 and 2. Two classification algorithms were trialled during testing - a multi-layer perceptron and an adapted k-nearest neighbour (KNN) algorithm (detailed below); with the KNN resulting in the better performance.

Receipt of a new state calls the prediction functions of the style and action-type learning algorithm instances. The resulting suggestions are then used to compose style and action suggestions. Social Supporting Action suggestions are composed through combination of the style and action-type suggestions according to the pairings in Table 4.2. If such a pairing does not exist, then no action suggestion is returned. The resultant output for one action-type and two different styles is shown in the final code excerpt presented in Algorithm 2. In line with the state analyser update rate, these prediction functions will be called at a frequency of 2 Hz, such that the Learner can be responsive to changes in e.g. user behaviour that require a response. However, the overall action suggestion rate is expected (and desired) to be much lower, such that many state inputs yield a return of no action suggestion.

Following Senft et al. (2019) on receipt of an expert-initiated/evaluated action, a reward value is generated according to its validation status (see Algorithm 1). Teacher-initiated actions and teacher-accepted learner-suggestions are given a reward of 1. Passively accepted suggestions are given a reward of 0. Learner-suggested actions that are refused by the expert are given a reward of -1. The *action-type* and/or *style* is then enumerated via a dictionary to create *{state-label-reward}* tuples which are added to the respective collections of instances representing classifier training data. Note that Task Actions generate style training data only, as the *action-type* is pre-set by the C25K programme. Use of these rewards in the context of making suggestions is shown in Algorithm 3.

4.2.5.6 Behaviour Manager

The Behaviour Manager (6) turns validated actions into explicit robot commands and executes them via the NAOqi ROS bridge. Dialogues to be used in speech-based actions are stored in dictionaries for each permissible *{action-type, style}* combination. To identify which specific dialogue to execute for a given action, the behaviour logs are checked to identify which dialogue was used last time the current user saw this action. The next listed dictionary entry is then selected, until all dictionary listings have been exhausted in which case the first dictionary entry is used and iteration through the dialogue begins again.

Algorithm 1 Processing of an Social Supporting Action training example.**Input:** Action e ($\{action, style\}, state$)

```
a = e.action
s = e.style
x = e.state
if a.validation == EXPERT_INITIATED or ACCEPTED then:
    r = 1
else if a.validation == PASSIVE_ACCEPTED then:
    r = 0
else if a.validation == REFUSED then:
    r = -1
end if
ca = (a,x,r)
cs = (s,x,r)
actiontype_learner.add_instance(ca)
style_learner.add_instance(cs)
```

Algorithm 2 Example creation of a Social Support Action suggestion on receipt of a new input state.**Input:** new state x:

```
suggested_style = style_learner.predict(x)
suggested_action = actiontype_learner.predict(x)
if suggested_style != NONE then
    create_update_style_action(style)
    if suggested_action != NONE then
        if action == SOCIAL then
            if style == POSITIVE then
                suggested_action = [HUMOUR]
                ...
            else if style == SYMPATHETIC then
                action = random_choice[SYMPATHISE, CHALLENGE]
                suggested_action = [SYMPATHETIC, action]
```

4.2.6 Supervised Learning: Information Flow

The following subsections describe example processing of a teacher-initiated and learner-suggested action respectively, to demonstrate the flow of information through the architecture during supervised operation of the system.

4.2.6.1 Teacher-Initiated Actions

1. The fitness instructor chooses an action via the Teacher Interface (1)
2. This is received by the Action Moderator (4) and

CHAPTER 4. CREATING AN AUTONOMOUS, SOCIALLY ASSISTIVE ROBOT WITH INTERACTIVE MACHINE LEARNING

Algorithm 3 Adapted KNN algorithm logic for generating suggestions; used for both style and specific action class (shown here as applied to style suggestions). Note that the *threshold* used to decide whether suggestions get proposed to the supervisor is dynamically updated as new tuples are added to the collection of instances (following Senft et al. (2019)).

Input: x' current state ; C_S collection of style instances $c_s = (s, x, r)$ S ensemble of styles present in C_S

Output: suggested style $\pi_S(x)$

for all $s \in S$ **do**

for all $p = (x, r) \in C_S$ **do**

 compute similarity Δ between x and x' : $\Delta(p) = 1 - \frac{\sum_{i=1}^n (x'(i) - x(i))^2}{n}$

 find closest pair $\hat{p} = \text{argmax}_p \Delta(p)$

 compute expected reward $\hat{r}(s)$ for applying s in state x' : $\hat{r}(s) = \Delta(\hat{p}) \cdot r(\hat{p})$ where $r(\hat{p})$ is the reward r of the pair $\hat{p} = (x, r)$

 select the style with the maximum expected reward: $\pi_S(x') = \text{argmax}_s \hat{r}(s)$

if $\hat{r}(\pi_S(x')) > \text{threshold}$ **then**

 propose $\pi_S(x')$ to supervisor

- a) published as a robot action to be executed via the Behaviour Manager (6) according to priority queue management
 - b) published as an action-type and/or style learning example.
3. Learning examples are received by the Learner (5) where they are paired with the latest input state received from the State Analyser (3), and a reward $r=1$ to create *state, action-type/style, reward* tuple(s) which are added to the relevant training dataset(s) as per Algorithm 2

4.2.6.2 Learner-Suggested Social Supporting Actions

1. The Learner (5) receives a new input state which is used to generate a suggested Social Supporting Action
2. This is received by the Action Moderator (4) and relayed to the Teaching Interface (1) according to priority queue management
3. The suggestion is displayed on the Teaching Interface (1) through highlighting of the relevant action icon. There are three ways in which the fitness instructor can respond, representing three validation states a validated action can take:
 - a) accept the action suggestion within 10s (accepted)
 - b) refuse the action suggestion within 10s (refused)
 - c) wait for the action suggestion to time out, in which case it will be considered *passively accepted*

4. The resultant *evaluated action* is fed back to the Action Moderator (4) and:
 - a) if accepted or passively accepted, published as a robot action to be executed via the Behaviour Manager (6) according to priority queue management
 - b) published as a style and/or action learning example, received and processed by the Learner (5) as previously described

4.3 Real World Deployment: Learning & Evaluation Study

An experimental protocol was developed to allow for training of the system, and then testing of the resultant autonomous behaviour, all within the context of delivering the fixed-term C25K programme. The protocol was also designed to allow for comparison of the IML system to a heuristic based alternative, but this is discussed in detail in Chapter 5. Additional observations regarding mutual shaping effects are also presented in that chapter. The research questions and results presented in this chapter are specifically concerned with investigating whether the approach taken was practical (as a process) and resulted in appropriate autonomous robot behaviour. The study was approved by the University of the West of England Research Ethics Committee.

4.3.1 Research Questions and Hypotheses

RQ1 How does the fitness instructor utilise and interact with the IML system?

H1A The co-designed action space and teaching interface will allow the fitness instructor to ensure an appropriate robot action policy during supervised sessions.

H1B Fitness instructor use/supervision of the system will result in personalised action policies for each participant.

H1C The IML system will reduce the fitness instructor's *active* workload over time: the number of Learner suggested actions which are accepted will increase and the amount of unprompted actions triggered by the instructor directly will decrease.

RQ2 Is the IML approach (and the amount of training data captured during initial phases of the study) sufficient to create appropriate autonomous robot behaviour?

H2A The autonomous robot will utilise the entirety of the co-designed action space in a similar way to the fitness instructor.

H2B The autonomous robot will demonstrate personalised behaviour across participants.

H2C Participants will not notice the switch from supervised to autonomous control of the robot, and will not evaluate the (autonomously running) robot significantly differently on post-session measures.

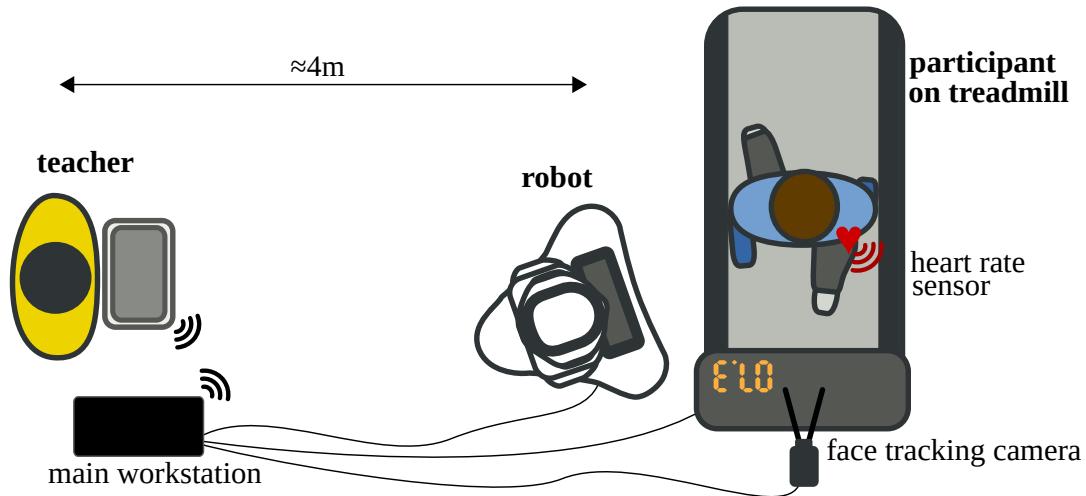


Figure 4.7: Overview of the experimental setup.

H2D The fitness instructor will evaluate the autonomous behaviour as being appropriate and effective.

RQ3 What is participants' experience of undertaking the Couch to 5km programme (with the C25k robot coach/fitness instructor) as per the experimental setup?

H3A Overall participant experience of the programme will be positive, with specific reference to the robot as a motivational aid.

H3B The (background) presence of the fitness instructor will be a factor in participants' acceptance/positive experience of working with the robot.

4.3.2 Gym Installation and Setup

The robot was installed in a university operated gym, on the University of the West of England's Frenchay campus, that was closed to any other users for the duration of the study. The experimental setup is depicted in Figures 4.7 and 4.8. The fitness instructor involved in co-designing the action/input space was also employed (paid at his normal hourly wage) to be the system 'teacher' and to observe/facilitate all sessions.

4.3.3 Participants

Participant recruitment was conducted through an advertisement and preliminary information sheet being shared via email (particularly via staff mailing lists to target those working on campus) and social media. Full details of the recruitment materials and pre-requisites are given in Appendix C. Importantly, whilst the advertisement indicated the study was about whether machine learning could be used to have a human expert train the robot, it did give any details on exactly *how* this machine learning setup would work, and the potential for varying autonomy was



Figure 4.8: Final experimental set-up; photograph of an exercise session undertaken during the study. Shows fitness instructor and robot position with regard to the participant, and the fitness instructor's use of the teaching interface to control/supervise the robot.

not mentioned. Health related inclusion criteria were relatively minimal given that the C25K programme is designed for people of all abilities, but the advertisements did attempt to target people who were somewhat interested in taking up running but did not run excessively already.

10 participants (4 male/6 female, age range 26 to 60 with mean 36.2) were recruited to take part in the study. One participant (female) dropped out midway through the study, their data were still used for training the system but have been excluded from detailed analysis of system use/performance. Participants were not required to be existing university gym members and were neither required to pay any sort of membership fee nor compensated for their participation in the study.

4.3.4 Fitness Instructor

The fitness instructor (male, 24) was recruited through the UWE Centre for Sport, whom also facilitated hosting the experiment with regards to gym access, insurance etc. He is a fully qualified instructor, employed by the Centre to lead group exercise sessions and offer personal training. He holds a CYQ Level 2 in Gym Instruction and Gym-Based Exercise and a YMCA Level 3 in Personal Training. He has worked in the fitness industry for 5 years, 3 of which have been at the UWE Centre for Sport. He also holds a BSc in Biological Science.

4.3.5 Conditions

Over the course of the in-the-wild testing, participants worked with the IML system and a heuristic based control, according to a specific testing schedule detailed in the following subsection. In total, participants saw three versions of the robot coach:

1. IML-Supervised (IML-S): The IML system as supervised/ultimately controlled by the fitness instructor, i.e. with him generating unprompted actions and responding to system generated action suggestions to generate training data.
2. IML-Autonomous (IML-A): The IML system allowed to run autonomously, i.e. with no additional actions generated nor suggested actions refused by the fitness instructor.
3. Heuristic: A heuristic based system, using heuristics derived during initial co-design and updated with the fitness instructor between study Phases 1 and 3 (Heuristic (v1) and (v2) per Table 4.7).

The Heuristic system was designed to represent the output of traditional participatory design approaches, as an alternative way to generate expert-informed behaviour autonomously. Like the IML system, the Heuristic system was also designed to generate social supporting actions to supplement the C25K programme-set Task Actions. The purpose of designing this system was act as a form of control condition, specifically to compare to the '*IML as participatory design*'

methodology proposed by this work. Further details about this system, and results comparing the resultant autonomous behaviours are therefore given in Chapter 5, where this methodology is presented and critiqued in full.

Importantly, whilst participants were made explicitly aware that they were testing out two different versions of the robot when seeing the IML-S and Heuristic systems, they were not given any indication as to *how* those robots were supposed to be different or how they were programmed/controlled differently. In contrast, the switch from IML-S to IML-A was done covertly with participants not being made aware of any change in the robot's control. Throughout the programme, no detail concerning the robot's learning process nor the actual supervisory mechanism concerning the fitness instructor's ultimate control of the robot was given to participants. As such, whilst participants were vaguely aware that the role of the instructor was to '*teach*' the robot as per the information sheet, they were never explicitly told how this was implemented. This is reflected in data collected on participants description of the instructor's role in the study/delivery of their exercise sessions, presented and discussed in Chapter 5.

4.3.6 Testing Schedule

The testing schedule was designed around delivery of the C25K Programme. Exposure to the three conditions/versions of the C25K coach was split across three key testing phases, as described below. Completed in full, with no breaks, the C25K programme takes 9 weeks to complete. Use of the gym was limited to a maximum of 12 weeks from when the study began, for which the robot was installed throughout, representing some flexibility for, but also a hard limit on, the ability to offer catch-up sessions in the case of participant/fitness instructor absence. As such, each participant's specific testing schedule was equivalent at the beginning (with the first 8 sessions alternating the IML and Heuristic robots) but updated dynamically toward the end of the study based on participant attendance and completion of sessions to ensure they all (i) completed one session with the updated Heuristic robot and (ii) completed at least two sessions with the IML system running autonomously. An example for participant LB, showing how each test phase related to specific C25K sessions and robot conditions, is given in Table 4.6. A full breakdown of how many sessions each participant completed, and which version of the robot they saw when, is given in Table 4.7. Training data was collected during all IML-S training sessions in Phases 1 and 2 of the study.

- Phase 1 [8 sessions per participant]: Participants alternated between the IML-S and H1 robots each session, for a total of 4 sessions with each. It was made explicitly clear that the two robots were programmed differently (each robot was colour coded either *orange* or *purple* using a simple paper neck collar) but not it was not explained *how* they were different. Condition ordering and colour labelling were randomly assigned to and counterbalanced across participants.

S#	Phase 1								Phase 2			Phase 3				
	1	2	3	4	5	6	7	8	9	...	22	23	24	25	26	27
Cdn	H1	IML-S	H	IML-S	H2	IML-S	H	IML-S		IML-S	IML-A	IML-A	H	IML-A	IML-A	IML-A

Table 4.6: Experimental testing schedule for participant LB, who completed all 27 Couch to 5km sessions. *S#* is the C25K session number, and *Cdn* is the robot condition.

- Phase 2 [9-13 sessions per participant]: Participants worked exclusively with the IML-S robot as it continued to be trained by the fitness instructor. The robot was still labelled and referred to as either the *purple* or *orange* accordingly.
- Phase 3 [3+ sessions per participant]: Participants *unknowingly* worked out with IML-A robot (i.e. they were not briefed about the change in robot control in any way) for two sessions before the updated H2 robot was explicitly re-introduced. Again, to hide the difference between the IML and H systems, participants were told (fictitiously) that while they'd been working with e.g. the *purple* robot the 'other half of the group' had been working with the *orange* robot, or vice versa, and now they were once again being given another opportunity to test and compare the two systems. Any remaining sessions after these were run with the IML-A robot.

The above highlights a key limitation resulting from supervised training *and* evaluation being conducted in the same study. Specifically, the sessions used to demonstrate and evaluate autonomous robot behaviour were quite late within the robot's deployment and C25K exercise programme, in sessions when participants might already be quite independent in their exercising. Additional studies with new participants might be utilised to consider how well the trained system would perform in earlier sessions of the programme, but the complexities arising from dynamic changes between participants and the robot, instructor and exercise programme itself as that progresses would make like-for-like comparison very difficult.

4.3.7 Experimental Measures

As per the input space of the learning system presented in Table 4.5, a range of data concerning participant performance and task engagement were collected as part of the training data. Continuous state data, including all raw sensory readings used to produce those 20 learner input features was logged at the same rate it was published to the system (2 Hz) with each state reading being given an unique identifier. Ideally, participant task engagement/performance would be compared across robot conditions to give objective measures of the impact of robot behaviour on participant behaviour. However, it was decided early on in the design of this study that such data would not form part of the formal experimental measures because:

- the complex testing schedule and nature of the C25K session progression makes it difficult to achieve controlled within-subject testing of the impact of robot conditions from session to

4.3. REAL WORLD DEPLOYMENT: LEARNING & EVALUATION STUDY

Participant	Supervised (total)	Heuristic (v1)	Heuristic (v2)	Autonomous	Total
LB	2,4,6,8,9*,10-22 (18)	1,3,5,7	25	23,24,26,27	27
FB	2,4,6,8-21 (17)	1,3,5,7	24	22,23,25,26	26
DB	1,3,5,7,9-18* (13)	2,4,6,8	21	19,20,22	22
JF	1,3,5,7,9-22 (18)	2,4,6,8	25	23,24,26,27	27
MR	2,4,6,8-18 (14)	1,3,5,7	21	19,20,22	22
DP	1,3,5,7,9-21,24 (18)	2,4,6,8	27	22,23,25,26	27
JW	1*,3*,5,7,9-13,15,16,20-22 (14)	2,4,6,8	25	23,24	25
GB	2,4,6,8-21,23 (18)	1,3,5,7	26	22*,24,25,27	27
PT	2,4,6,8-18 (14)	1,3,5	21	19,20,22	22
MB	1,3*,5,,9-10,12-14 (7)	2,4,6,8	-	-	14
Total:	151	40	9	32	232

Table 4.7: Total number of experimental sessions conducted broken down per participant and experimental condition/robot automation strategy. Sessions marked* were subject to some technical issue(s) affecting robot performance and/or data collection. Note that participant MB dropped out after 14 sessions, and so results from their sessions are no included in the following analyses.

session, with session to session variation likely to be significantly impacted by factors like increasing session difficulty and increasing participant fitness etc.

- the small number of participants would reduce the extent to which any formal statistical analysis would have meaning.

All actions suggested by the Learner (along with the fitness instructor's response) or triggered by the fitness instructor were logged throughout the study. State data was also captured and stored at the system frequency of 2Hz. Each action log also therefore included the unique identifier of the corresponding record of input state data (i) that triggered it (for Learner suggestions) or (ii) was captured at the time of action selection by the fitness instructor (in the case of unprompted actions). These logs were designed to allow for (i) post hoc simulations of sessions, (ii) post hoc application of alternative machine learning methods/investigations regarding fitness instructor action/style choices and (iii) allow for objective analysis of system performance.

Explicit experimental measures, designed to capture participant experience of the programme and various robot conditions (as well as to detect whether participants noticed the switch from supervised to autonomous operation of the IML system) were as follows. Copies of all questionnaires, as administered to participants, are given in Appendix C:

1. A brief post-session questionnaire to be completed after every session: participants were asked how they found the session and how they would rate the robot as a fitness instructor. Included a 3-point emoji-based response (see Figure 4.9) as well as an open textbox.
2. A weekly 'journal' designed to be filled in in more depth with regard to their experience of the programme and working with the robot

3. A detailed questionnaire comparing the IML-S and Heuristic robots within-subject at the end of Phase 1 testing. Results for this are presented in Chapter 5.
4. Another questionnaire very similar to (3) at the end of the study, also discussed further in Chapter 5. However, pertinent to the work in this chapter, this questionnaire included questions on whether participants' perceived any change in either of the robot's behaviour over the course of the study. Regarding the overall study, participants were also asked to discuss the role of the robot versus the fitness instructor, and for their overall thoughts on using a robot for this purposes.

In all questionnaires, the two different versions of the robot (IML-S/A versus Heuristic) were referred to as the Orange or Purple robots according to the participants' colour convention allocation. As a reminder: participants were never told about the differences in how these robots were programmed/controlled, and were further told that whichever robot they had been working out with, the 'other half of the group' have been working out with the other one (to minimise the chance of participants suspecting that one robot was gradually learning/being 'taught' whilst the other was not).

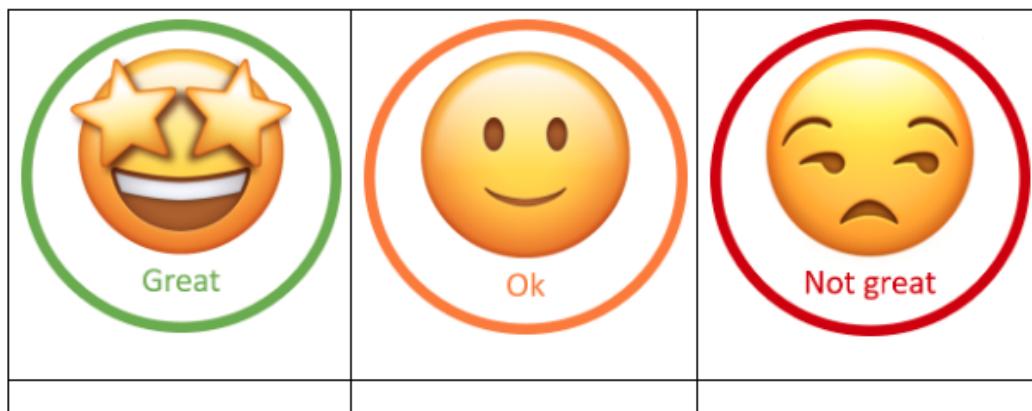


Figure 4.9: Emoji-based (Great / Ok / Not great) scale used in end-of-session participant and fitness instructor measures as well as for participant response to the C25K robot's Check PRE action.

The fitness instructor was similarly asked to complete a post-session questionnaire similar to (1) but this proved to be unmanageable as the study progressed and the free time between back-to-back participant sessions reduced. Instead, he kept one overall journal of unprompted notes regarding both system and participant performance that he captured in real-time alongside supervising sessions.

4.3.8 Technical Limitations and Run-time Fixes

Facial Expression Input

As the study progressed and participants were running faster, for longer, it became clear that the facial expression tracking started to fail; likely due to vibrations from the treadmill causing the camera image to blur. As such, the two facial expression features listed in Table 4.5 were removed from the input vector and the learning agent was re-trained based on all previous training data with those two features removed.

Rate of Action Suggestion

As noted previously, successful automation would require the learning system to identify *what* actions, *when* and *for who*. Preliminary testing of autonomous behaviour produced by the system (ahead of Phase 3) suggested a failure to learn one key element of the *when*; that sometimes it is actually most appropriate to do *nothing*. Dynamic updating of the policy suggestion threshold failed to impact the rate of suggestions such that suggestions were made every time the learner was passed an input state. This resulted in action suggestions being much too frequent and/or repetitive.

A ‘safety limit’ constraining suggestion rate to 1 action suggestion every 10s had already been coded within the Action Moderator to prevent the teacher tablet being rushed with suggestions, as experienced during early testing of the supervised system. So, for Phase 3 of the study, after discussions with the fitness instructor, this limit was increased to 30s. This matched the action rate of the hard-coded Heuristic system, also resulting from earlier iterative testing with the fitness instructor, with the motivation being that it would allow fairer testing of participant experience of the action choices made by the learner when running in autonomous mode.

Integration of Cooldown into Robot Led Session

Initially, the programme database was set to contain only the walk/run segments, detailed by the C25K plan, with a 5 minute warm up walk but no cooldown. It was assumed that participants leaving the gym and walking back across campus and would suffice, and that the fitness instructor would recommend and demonstrate some appropriate stretches. However as the programme progressed and participants were being asked to run for 20 minutes or longer with no breaks, the fitness instructor felt that a 3 minute cool-down walk ought also to be added to the mandated programme.

This was only implemented for sessions 18 onwards and only from week 7 (out of 12) of the robot being installed in the gym. For some participants, the transition to autonomous robot behaviour was made as early as session 19, and at the start of week 7 participants could already be up to session 21, so there was relatively limited training data available for this cooldown part of each session. However, this does make for an interesting feature to study when considering the Learner’s needs for training data, and potential to adapt to changes in the task, given its late addition to the programme.

Participant	Expert Initiated (r = 1)	Instant Accept (r = 1)	Passive Accept (r = 0)	Rejected (r = -1)	Total Training Data
LB	736	132	70	603	1541
FB	615	125	68	530	1338
DB	617	106	63	248	1034
JF	773	130	59	603	1565
MR	715	147	58	511	1431
DP	874	78	74	390	1416
JW	613	85	41	417	1156
GB	730	111	59	503	1403
PT	684	100	83	457	1324
MB	425	68	39	257	789
Total:	6782	1082	614	4519	12997

Table 4.8: Total number of training data points (*state-label-reward* tuples) collected per participant during supervised training sessions, broken down by validation type/reward value).

4.4 Findings

As summarised in Table 4.7, through the course of the study a total of 232 robot-led exercise sessions were conducted, of which 32 were run autonomously using (exclusively) actions suggested by the IML Learner, based on the training data collected during 151 supervised sessions. Table 4.8 lists the amount and type of training action examples (i.e. whether they were initiated by the learner then accepted/refused or executed unprompted by the fitness instructor) collected during the supervised IML-S sessions, broken down by participant. Details regarding system behaviour and experimental results in these sessions is given below. Details of the Heuristic sessions (in which the robot also operated autonomously) are presented in Chapter 5, as is more detail on mutual shaping effects observed concerning fitness instructor-robot-participant interactions throughout the study.

4.4.1 Example Session

Table 4.9 lists all the training action examples generated in an example supervised session (participant LB session 22). Whilst a single session cannot offer much insight on overall system performance, it does give a snapshot example of how the fitness instructor was using the system towards the end of the training sessions, and the real-time interaction between him and the Learner. In this session, it can be seen that the instructor accepts quite a few of the suggestions made by the Learner earlier on in the session, and doesn't seem to have to supplement these with many additional unprompted actions. This does not hold for the second half of the session however, where lots of refused suggestions and unprompted actions can be seen. The Δt value, which shows how long between the current action/suggestion and the last *non-refused* (i.e. fitness instructor unprompted action or accepted Learner suggestion), demonstrates the variability in rate of action execution. Two key observations can be made:

- (i) There are numerous examples of the fitness instructor executing unprompted actions very quickly (<10s) after another unprompted action/accepted suggestion. From the participants'

4.4. FINDINGS

Time (s)	Action ($\Delta t(s)$ since last <i>non-refused</i> action)	Time (s)	Action ($\Delta t(s)$ since last <i>non-refused</i> action)
11	learner suggests positive praise - PA	1,111	learner suggests positive praise - R (1)
48	learner suggests sympathetic challenge - A (37)	1,151	learner suggests positive praise - R (42)
84	learner suggests sympathetic sympathise - R (36)	1,187	learner suggests positive praise - R (78)
99	instructor initiates positive maintain (51)	1,225	instructor initiates positive praise (116)
122	learner suggests positive maintain - R (23)	1,226	learner suggests positive maintain - A (1)
160	learner suggests positive praise - A (61)	1,262	learner suggests positive maintain - R (36)
169	instructor initiates positive maintain (9)	1,297	learner suggests positive praise (71)
200	learner suggests positive maintain - A (31)	1,334	learner suggests positive praise - R (108)
237	learner suggests positive maintain - R (37)	1,372	learner suggests positive praise - R (146)
274	learner suggests positive praise - A (74)	1,410	learner suggests positive praise - A (184)
283	instructor initiates sympathetic sympathise (9)	1,431	instructor initiates challenging challenge (21)
313	learner suggests sympathetic challenge - A (30)	1,443	learner suggests challenging challenge - R (12)
352	learner suggests sympathetic sympathise - R (39)	1,448	instructor initiates sympathetic challenge (17)
391	learner suggests challenging challenge - A (78)	1,477	learner suggests challenging challenge - R (29)
431	learner suggests challenging challenge - R (40)	1,506	instructor initiates get_closer (58)
470	learner suggests challenging challenge - A (79)	1,508	instructor initiates sympathetic time (2)
508	learner suggests challenging challenge - R (38)	1,513	learner suggests get_closer - A (5)
541	learner suggests challenging challenge - A (71)	1,518	instructor initiates challenging time (5)
576	instructor initiates positive praise (35)	1,546	learner suggests sympathetic time - R (28)
580	learner suggests challenging challenge - A (12)	1,560	instructor initiates challenging challenge (42)
617	learner suggests challenging challenge - R (37)	1,565	instructor initiates challenging time (5)
651	learner suggests positive praise - A (71)	1,584	instructor initiates challenging performance (19)
655	instructor initiates sympathetic challenge (4)	1,585	learner suggests challenging time - A (1)
684	learner suggests positive praise - R (29)	1,603	instructor initiates positive praise (18)
724	learner suggests positive praise - A (69)	1,620	instructor initiates positive social (humour) (17)
760	learner suggests positive social (humour) - A (36)	1,625	learner suggests challenging time - R (5)
800	learner suggests positive praise - R (40)	1,634	instructor initiates sympathetic time (14)
840	learner suggests positive praise - A (80)	1,651	instructor initiates positive social (humour) (17)
843	instructor initiates positive time (3)	1,657	learner suggests positive social (humour) - R (6)
875	learner suggests positive praise - R (32)	1,678	learner suggests positive style for cool down walk - A (21)
899	instructor initiates challenging challenge (24)	1,711	instructor initiates sympathetic praise (33)
907	instructor initiate sympathetic challenge (8)	1,713	learner suggests positive animation - A (2)
909	learner suggests positive time - R (2)	1,724	instructor initiates check_pre (11)
944	learner suggests positive praise - R (37)	1,736	instructor initiates sympathetic praise (12)
978	learner suggests challenging challenge - R (71)	1,747	learner suggests sympathetic praise - R (11)
1,011	learner suggests challenging challenge - R (104)	1,780	learner suggests sympathetic praise - R (44)
1,047	learner suggests get_closer - R (140)	1,817	learner suggests check_pre - R (81)
1,080	learner suggests get_closer - R (173)	1,856	learner suggests sympathetic praise - R (120)
1,082	instructor initiates positive praise (175)	1,896	learner suggests sympathetic praise - PA (160)
1,100	instructor initiates positive maintain (18)	1,929	learner suggests sympathetic praise - R (193)

Table 4.9: Example supervised session: full list of actions either suggested by the system (alongside the instructor's response coloured red -R for refused, orange -PA for passively accepted and green -A for instantly accepted) or triggered by the instructor directly (highlighted in black, bold font) during LB's final IML-S (supervised) session. In the case of successful learning, it would be expected that very few actions would need to be generated or refused by the fitness instructor. Style updates (affecting robot eye colour) suggested by the system and automatically executed according to the queuing system described in Section 4.2 are excluded for clarity.

point of view, these ‘double-action’ combinations essentially resulted in a single action (with a longer stream of speech) from the robot. This appeared to be something the fitness instructor did repeatedly, intentionally, towards the end of the study. Based on his feedback, this may have been a result of his perception that the range of dialogue had been over-used and become repetitive/boring by this point.

- (ii) There are segments from approx 900s onwards where there were lots of action refusals but these were not ‘replaced’ by unprompted actions, i.e. few actions were executed at all (high Δt). This was explicitly highlighted by the fitness instructor when reflecting on one of the difficulties in supporting long, pure runs in comparison to the sessions composed of shorter, alternate walking and running sessions. Specifically he suggested that on longer runs it is sometimes best to let people ‘zone out’ and not to interrupt/distract if they were ‘in the zone’.

This complex use of actions/action timing likely contributes to why the Learner failed to produce an appropriate rate of action suggestion, as referred to in Section 4.3.8. Further considering (ii) as refused suggestions were not always replaced with unprompted actions, it wasn’t necessarily true that the Learner’s suggestions were of the incorrect *style* or *type*. Rather, it was simply more appropriate to do nothing at all. The concept of doing *nothing* as sometimes being the best course of action represents something that the Learner fundamentally failed to replicate, likely due to a fundamental flaw in the (i) dynamic threshold approach to limiting action suggestions and/or (ii) the lack of encoding ‘*do nothing*’ as an actual ‘action’ available to the system. In addition, given that only sessions 18+ were pure runs of this kind, only a relatively small amount of the training actions/sessions were of this type.

However, Table 4.9 does show evidence for successful learning of a specific *what* action *when* pairing. As discussed under Section 4.3.8, the cooldown walk was only integrated into the system quite late on during the study, and LB was one of the participants furthest ahead with programme progression by Session 18 (such that he was one of the first participants to be exposed to this new cooldown section). As such, a high level of inappropriate/refused suggestions might be expected during this period. However, the Learner did suggest an *animation* action at $t = 1713\text{s}$ which was accepted by the fitness instructor. This very much reflects the fitness instructor’s use of the animation action almost exclusively within the cooldown period of the pure run sessions.

4.4.2 Usability for Generating Appropriate Action Policies (RQ1)

These findings are presented to ascertain whether the IML system design and implementation, including the robot actions, the teaching interface and the Learner suggestion - fitness instructor validation pipeline, allowed the fitness instructor to control the robot as desired. Specifically, through executing unprompted actions and responding to Learner suggestions, the instructor

should have been able to ensure an appropriate action policy was executed during all supervised sessions.

4.4.2.1 Instructor Use of the IML System

Figure 4.10 shows the breadth and relative use of action/style pairings executed during the supervised sessions (hence also representing the training data fed to the Learner). Note than *run* and *walk* actions are included. Whilst these actions were fundamentally set by the C25K programme their style could be set by the fitness instructor/Learner. It can be seen that whilst some actions were utilised much more than others, the fitness instructor made use of the entire robot action space available to the IML system. This demonstrates the utility of the co-designed Social Support actions.

The example session presented in Table 4.9, as well as the cumulative plot of unprompted actions and accepted/refused Learner suggestions in Figure 4.17, demonstrate that the fitness instructor was able to actively manage the Learner's suggestions (accepting appropriate actions and refusing inappropriate actions) whilst still executing his own unprompted actions where necessary. As such, the instructor ultimately retained control of the robot's behaviour throughout the supervised sessions. Three key observations suggest that through this control, the instructor was able to generate an appropriate action policy. Firstly, as discussed under Section 4.4.1, the instructor described the need to utilise different behaviours across the sessions that were made up of alternate running and walking versus pure running. Figures 4.11 and 4.12 show the variation in (i) relative action type/style use and (ii) number of actions executed (normalised against session length) for these two types of supervised sessions.

Secondly, the adjustment of action policy was also demonstrated in the context of personalisation, with evidence that different actions/styles were utilised differently across participants during supervised sessions. This is discussed in more detail in the following subsection. In addition, post-session participant ratings of the robot as a fitness instructor were overwhelmingly positive (as can be seen in Figure 4.26), suggesting the executed actions were effective and appropriate. Finally, in a final post-study interview in which the instructor was asked how he found the training process and whether he found the system intuitive to use, he stated:

"It was fairly smooth... and because we had designed it together, I knew exactly what I wanted, where it was. And my sort of navigation around the system. So I mean, it made my implementation of what I wanted pretty smooth."

These results provide strong support for H1A: *the co-designed action space and teaching interface will allow the fitness instructor to ensure an appropriate robot action policy during supervised sessions*. This likely results from the carefully considered activities of the co-design process (as outlined in Table 4.1).

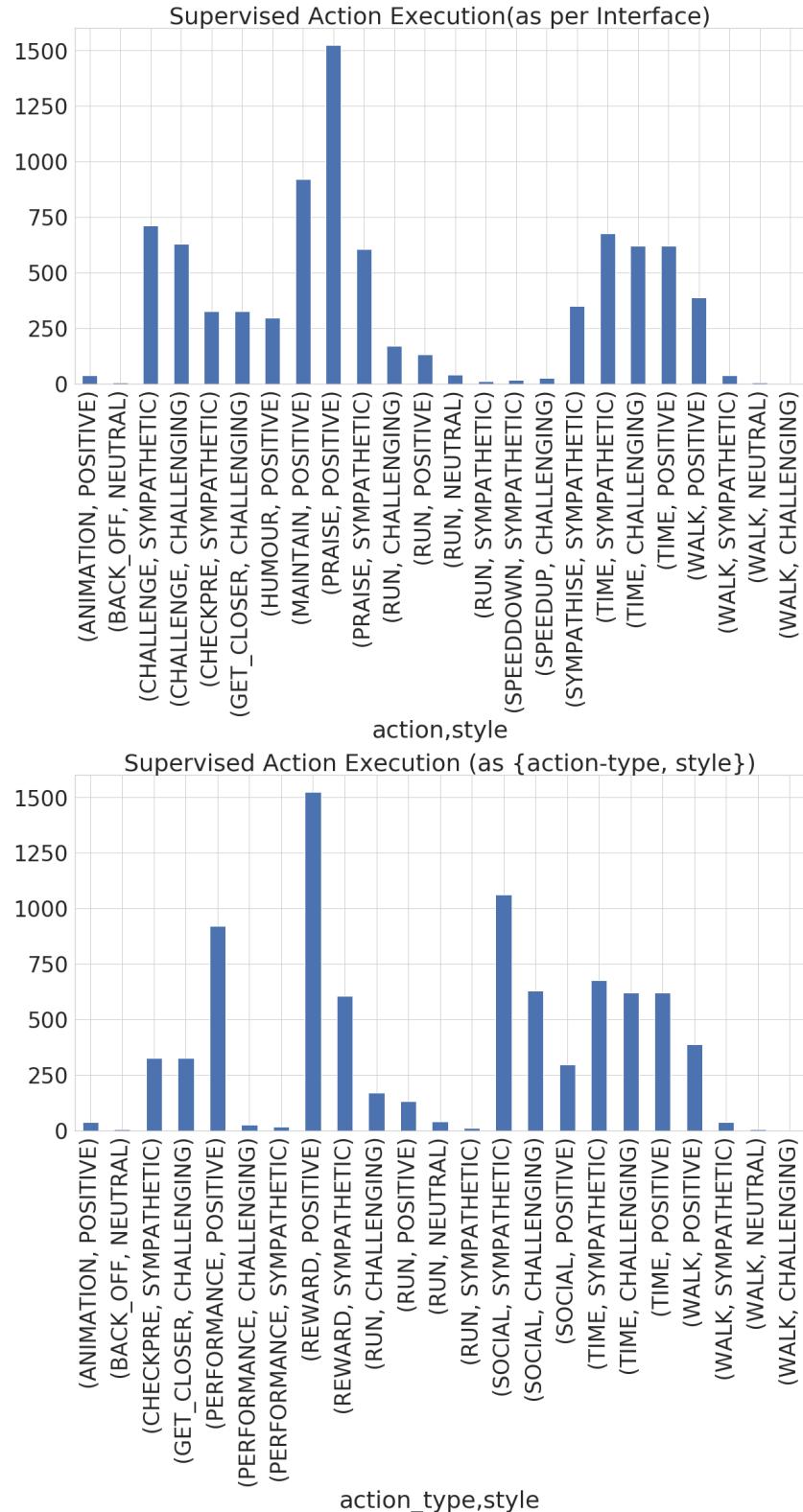


Figure 4.10: Actions (and their style) executed during sessions supervised by the fitness instructor. Grouped by (i) actions and styles as represented on the teaching interface (shown in 4.4) and (ii) (*action-type, style*) pairings as per the abstraction of the instructor-designed actions presented in Table 4.2.

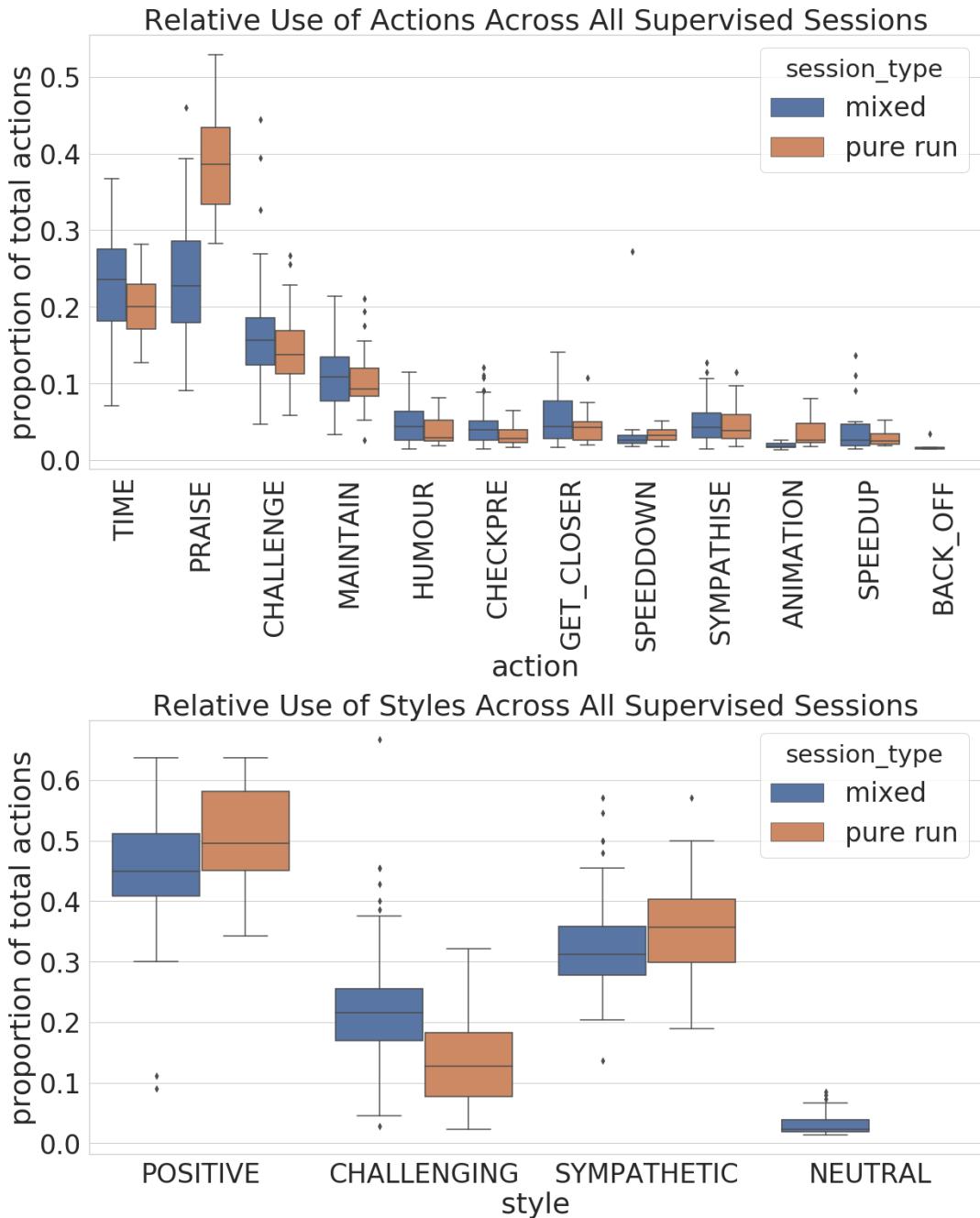


Figure 4.11: The relative use of each action and style across mixed versus pure run sessions supervised/ultimately controlled by the fitness instructor. Whilst the differences may not be statistically significant, they suggest the instructor adjusted the action policy, via supervised control of the IML system, to account for the change in session type.

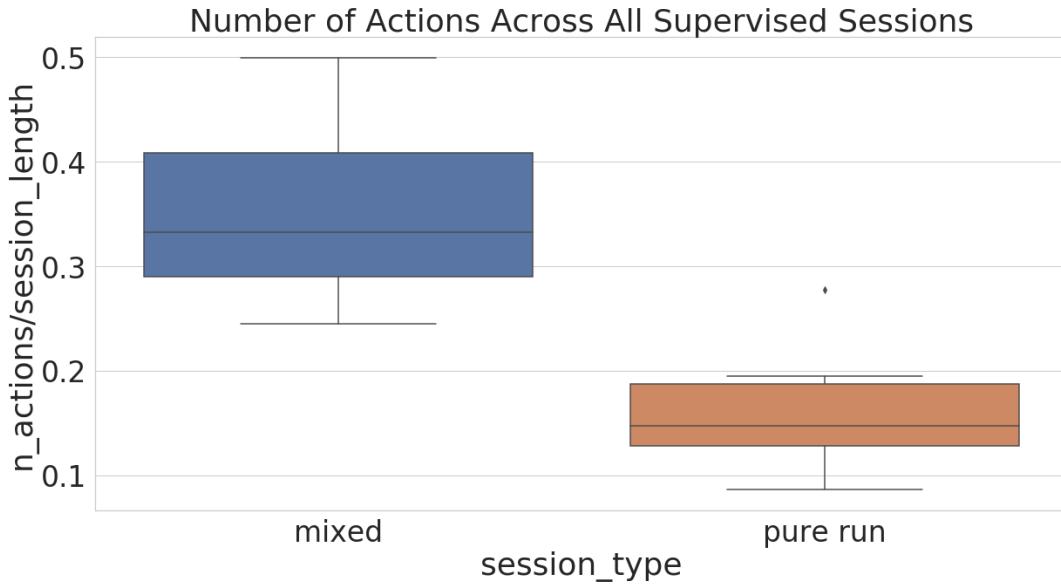


Figure 4.12: The number of actions used in mixed versus pure run sessions supervised/ultimately controlled by the fitness instructor, normalised with respect to session length (in seconds). Note that less actions were typically executed in the pure run sessions, in line with comments made by the fitness instructor. This gives an objective demonstration of an action policy change described qualitatively by the instructor and evidenced objectively in his supervised output of the system.

4.4.2.2 Personalisation

To investigate personalisation by the fitness instructor, data was compared across participants for the last, common, supervised session of the programme that they all completed - session 17. Figures 4.13 and 4.14 show how style and action-type of actions executed varied across each participant's session 17. As an alternative way of demonstrating this variation, Figure 4.15 shows how full actions (combining *action-type* and *style*) executed in *each* participants' session 17 contributed to the cumulative sum of actions across *all* participants' session 17. If the action/styles executed were very similar across all participants, then each participant should contribute approximately 1/9th of the cumulative sum of each action/style combination. Clearly this is not the case, e.g. with some actions only being used for a single participant.

Another element of personalisation that these figures fail to show is *when* what types of actions were executed. For example, when the fitness instructor was asked to identify participants he felt required very different training approaches, he identified LB vs MR, FB vs JF and GB vs PT. The action distributions for MR and LB in Figure 4.14 don't look significantly different. However, Figure 4.16 shows there are clear differences in *when* those actions were executed within the session. Together, these results provide strong support for H1B: *fitness instructor use/supervision of the system will result in personalised action policies for each participant*. Results on the importance of personalisation from Chapters 2 and 3 would suggest this personalisation would be driven by the fitness instructor (i) responding to participants' instantaneous

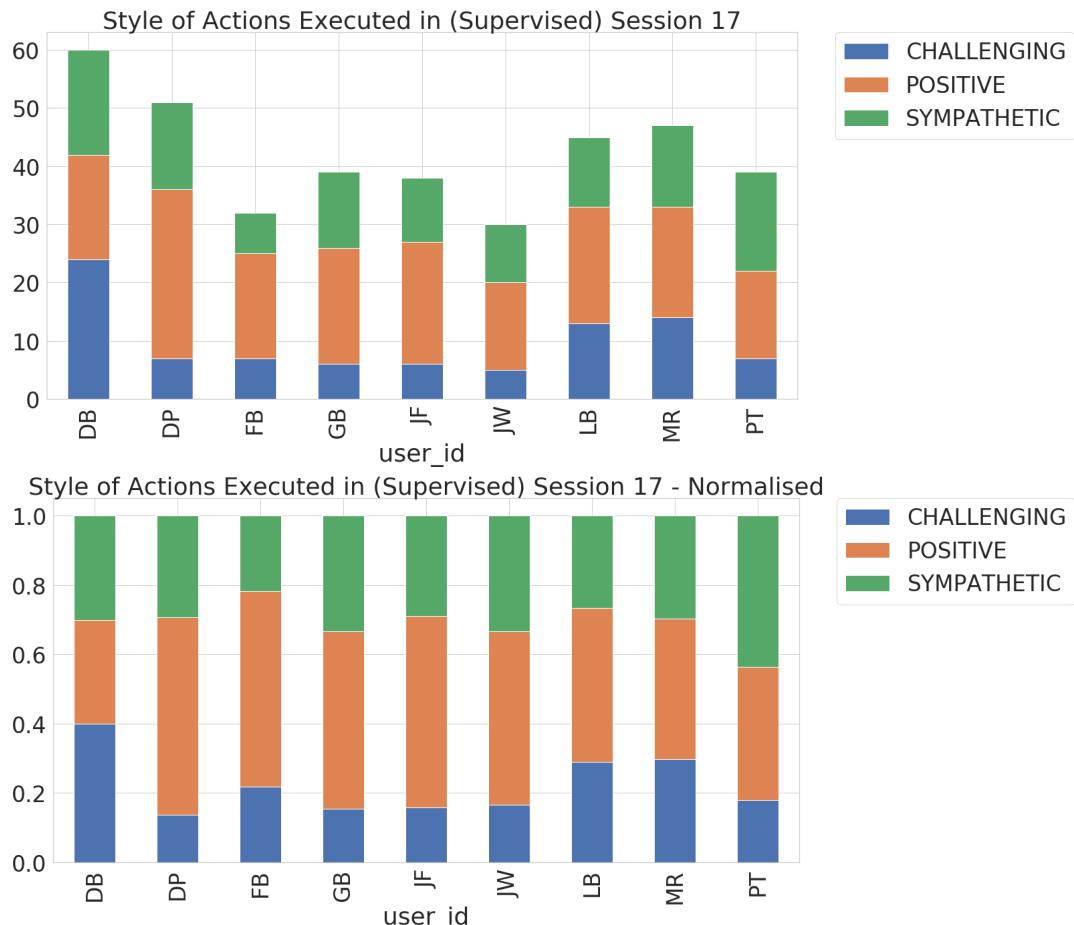


Figure 4.13: The style of executed actions across participants in the final, common, supervised session that they all completed. Plotted firstly as raw data (showing the overall number of actions varied across participants) and then normalised, demonstrating the instructor adjusted the styling of actions across participants. If robot behaviour was informed only by programme progress and not individual performance and personality, then we would expect (i) the number of actions and (ii) the proportion of action styles to be approximately equal across participants.

performance/engagement and (ii) shaping robot behaviours based on his overall knowledge of the participant and his perception of their relationship with the robot. The latter is particularly evidenced by the instructor's in-session notes, which often included comments on participants' apparent engagement with the robot and/or whether e.g. they needed to be challenged more or were already pushing themselves very hard.

4.4.2.3 Instructor Workload

Figure 4.17 plots the data described in Table 4.8 to show the progressive accumulation of unprompted actions, accepted Learner suggestions and refused Learner suggestions. This essentially demonstrates how the fitness instructor's interactions with the Learner developed over the course

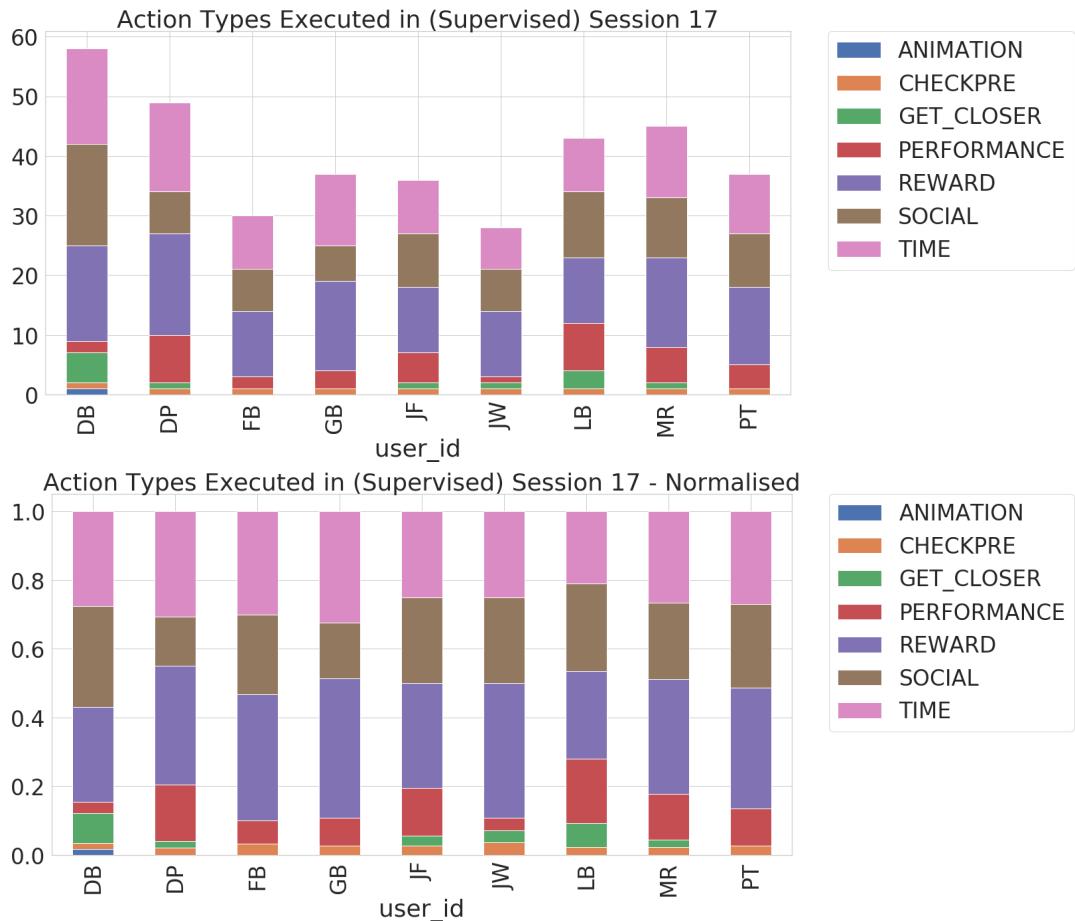


Figure 4.14: The action-type of executed actions across participants in the final, common, supervised session that they all completed. Plotted firstly as raw data (showing the overall number of actions varied across participants) and then normalised, demonstrating the instructor adjusted utilisation of the action-types across participants. If robot behaviour was informed only by programme progress and not individual performance and personality, then we would expect (i) the number of actions and (ii) the proportion of action types to be approximately equal across participants.

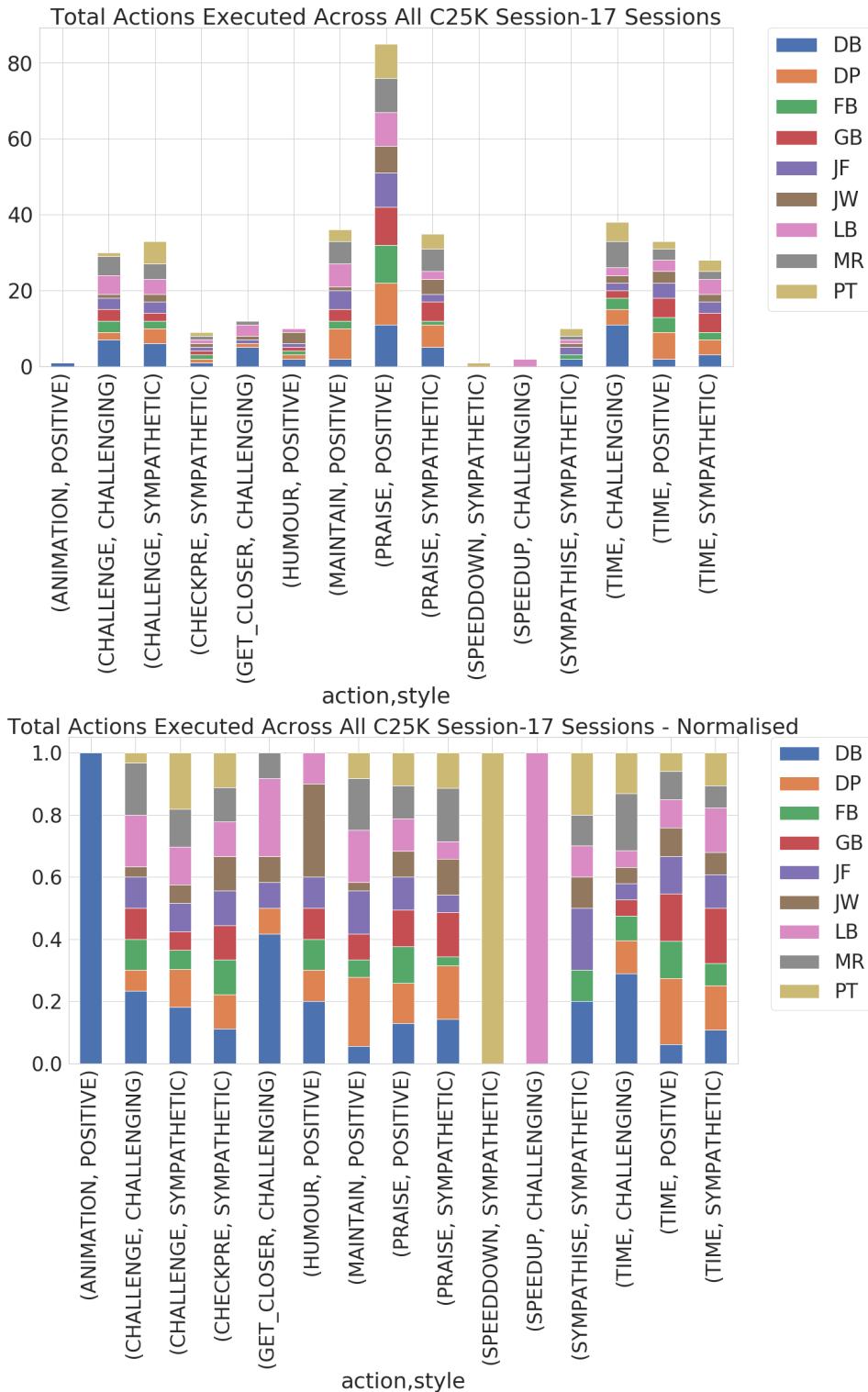


Figure 4.15: The total of each action (and associated style) executed across all participants' session 17, broken down by participant. The normalised version particularly shows the variation across participants; if the action distribution was the same across participants then we would expect each participant to contribute 1/9 of each bar.

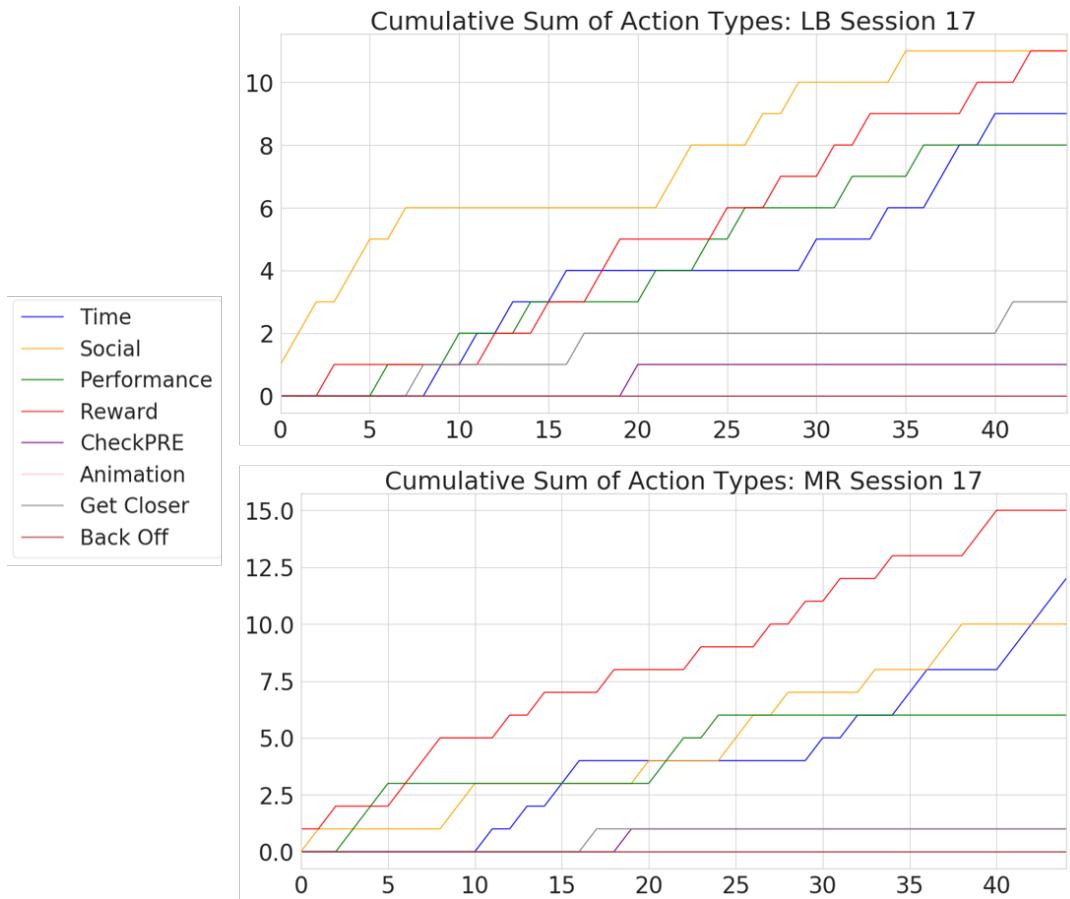


Figure 4.16: The cumulative sum of each action type throughout LB and MR's (supervised) session 17; showing *when* each action type was used through the session to demonstrate this additional element of personalisation in the fitness instructor's supervised execution of actions. Whilst the total instances of each action type isn't that different between the two (as shown in Figure 4.14) it can be seen that there are clear differences in *when* those actions were executed within the session. For example, lots of *social* actions were used at the start of LB's session, whereas *reward* actions were used at the start of MR's session. Similarly, *performance* actions were utilised much earlier in MR's session than in LB's.

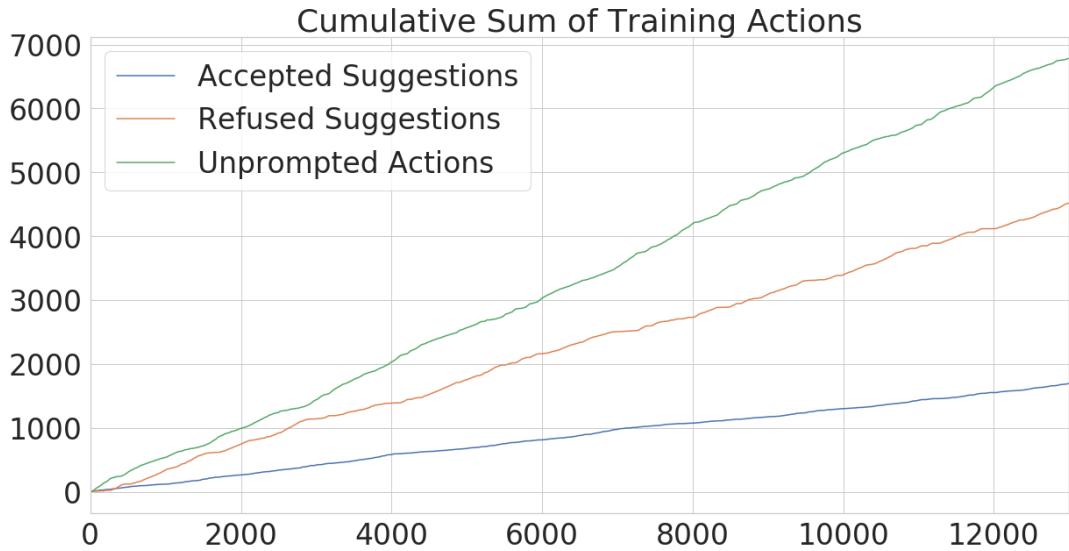


Figure 4.17: Cumulative sum of actions (accepted suggestions, refused suggestions and unprompted exemplars) collected as training data, showing that there was little change in the rate at which the expert was producing unprompted exemplars or accepting/refusing action suggested by the learner. For successful learning and the related expected expert workload reduction, we would expect the number of refused suggestions and unprompted actions to plateau as the number of accepted suggestions continued to rise.

of the supervised sessions, up to the end of Phase 2 testing. Whilst there was a steady increase in the number of accepted Learner suggestions, notably the graph shows no suggestion of a decreasing rate in refusals of Learner suggested actions, nor a definite reduction in the rate of unprompted actions.

In a given session, the overall ‘active’ workload of the fitness instructor is made up of the need to (i) refuse inappropriate Learner suggestions (including *style* suggestions applied to fixed task actions (i.e. the instructions to switch between running and walking) and (ii) generate additional unprompted Social Supporting actions if required. Given the dual-learner setup of the system, whereby task actions are reliant only on the output of the style learner, it is useful to consider the Task and Social Supporting actions separately. Figure 4.18 plots the instructor’s response to Learner suggested Task Action styling as training progressed, i.e. plotted for each session of Phase 2 in chronological order. It can be seen that quite quickly, the instructor stopped needing to overrule the Learner’s suggested style at all. This suggests the Learner was very successful in learning what *style* or *mood* the robot should be in.

Figure 4.19 similarly shows how overall workload varied as training progressed, within the context of the overall number of actions suggested by the Learner and executed unprompted by the instructor. To support H1C: *the IML system will reduce the fitness instructor’s active workload over time*, we would expect the number of accepted actions to increase whilst the number of refusals and unprompted actions decreased. However, in the figure it can be seen that by the

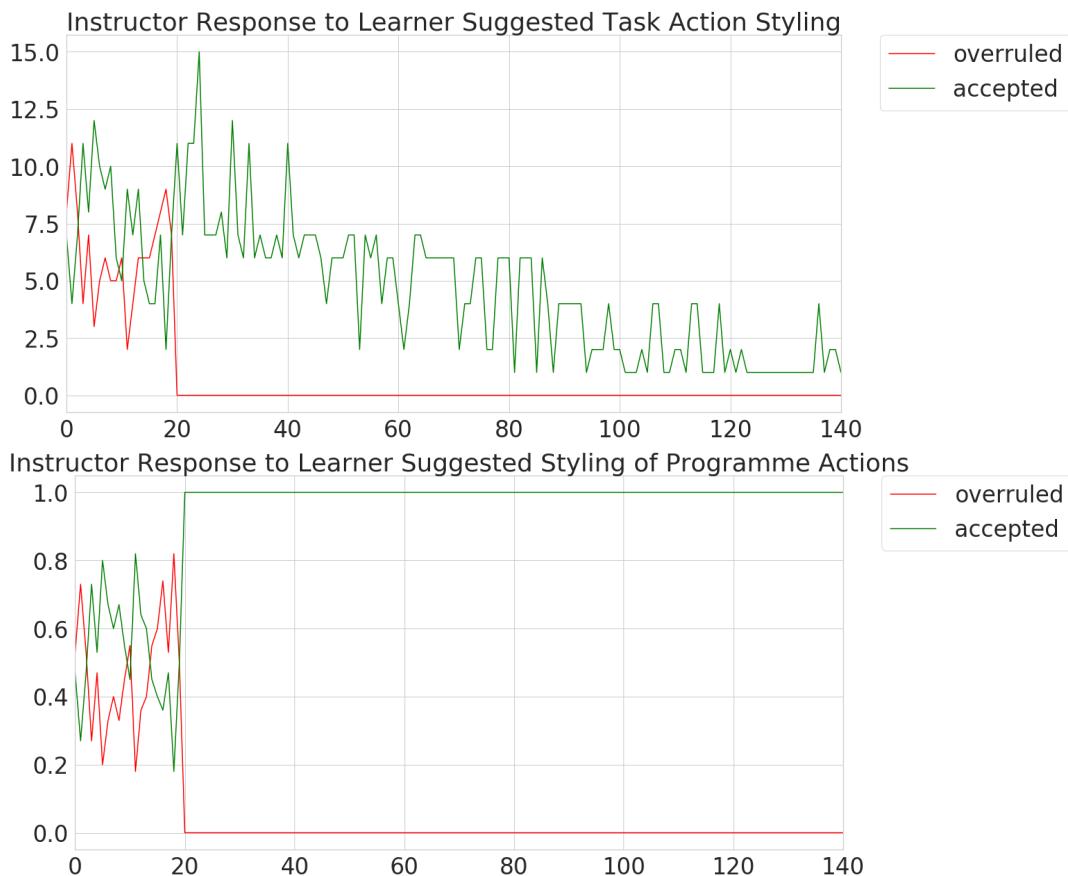


Figure 4.18: Plot of instructor response to Learner suggested task action styling across all supervised sessions of Phase 2. Raw action numbers plotted first to highlight the actual number of successful style suggestions. Noise in raw actions numbers is due to participants simultaneously being at differing stages of the programme (hence experiencing a different number of task actions). Normalised data also plotted for clarity. From approximately the 20th study session onwards, the instructor accepted all Task Action style suggestions made by the Learner. Depending on progress through the programme, this represented between 1 and 15 successful action style suggestions per session.

end of Phase 2, the instructor was still very actively refusing Learner suggestions - across all participants' final (common) supervised session (session 17) 72.4% of learner suggestions were rejected and 27.6% were accepted. Similarly whilst the number of unprompted actions required did seem to be generally decreasing, they still represented the majority of all executed actions. Across that same final common supervised session, 76.0% of actions were generated by the instructor with only 24% being accepted Learner suggestions. As such, there is very little support for H1C. However, the success of suggested *styles* for the task (run/walk) actions demonstrates the potential for this element of the system to be real world useful.

The most obvious explanation for the low acceptance of Learner suggestions would be that the Learner was simply suggesting a significant number of inappropriate actions. Rate of action suggestion has already been discussed as something the system failed to properly achieve, such that the sheer number of action suggestions (appropriate or not) was likely one of the key reasons for this high rejection rate. Further, as discussed in the following section, the resultant autonomous behaviour (and its evaluation) actually suggests that the Learner suggested actions were not inappropriate with regards to content, but were indeed potentially too frequent and/or therefore repetitive.

Another factor that may contribute to a lack of workload reduction is simply that the instructor did not *allow* the workload to reduce, i.e. he (subconsciously or otherwise) actively wished to remain in control of/guiding the session and therefore possibly had a tendency to take an active role in rejecting suggestions in favour of slightly different actions or a slightly different timing of the same action. This would be in line with similar results documented by Senft et al. (2019) who also found that the expert in the loop was still taking a very active role in managing supervised robot behaviour and rejecting Learner suggestions towards the end of training, even when autonomous testing then suggested that Learner generated, autonomous behaviour was fairly appropriate and similar to the experts' supervised behaviour. Future work might investigate this further in part by considering expert instruction, training, supervision and 'handing over of responsibility' to an equivalent *human* trainee.

4.4.3 Autonomous Robot Behaviour (RQ2)

These findings are presented to ascertain whether, when running autonomously, the robot successfully behaved appropriately in supporting participants through the programme.

4.4.3.1 Comparison to Supervised Behaviour

The nature of the C25K interaction scenario makes it difficult to compare action distributions across conditions as a measure of performance, as exercise sessions are incredibly dynamic with regards to e.g. participant state (energy level that day, mood, fatigue etc.) and task requirements, e.g. lots of short run/walks versus longer runs. As such, two 'good' sessions, where the robot acts appropriately, may have very different action distributions. However, comparing the overall

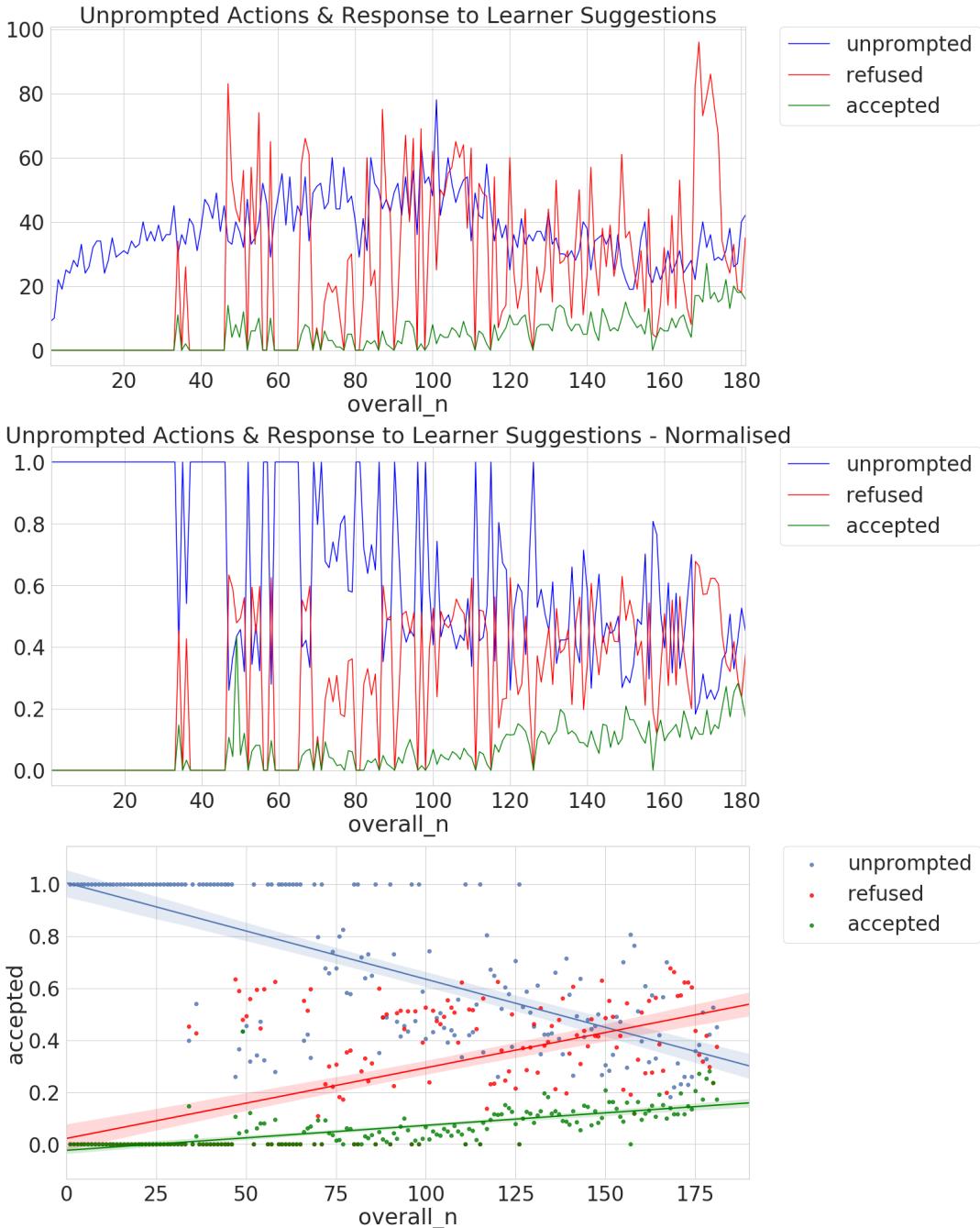


Figure 4.19: Summary of the fitness instructor's workload, concerning the generation of unprompted actions and responding to Learner suggestions, and how that evolved over the supervised sessions of Phase 2. Raw action numbers plotted first to highlight the actual number of actions occurring during sessions. Normalised data also plotted for clarity, as well as a scatter with line of best fit to show any overall trend. Whilst an increase in accepted Learner suggestions can be seen (as would be expected if the Learner was getting 'better' based on the increasing amount of training data collected) there is not an overall decrease in workload (*active* or *supervisory*) as the instructor still generated a lot of unprompted actions and refused a lot of the Learner's suggestions.

Participant	Supervised	Autonomous
LB	21, 22	23, 24
FB	20, 21	22, 23
DB	16, 17	19, 20
JF	21, 22	23, 24
MR	17, 18	19, 20
DP	21, 24	25, 26
JW	21, 22	23, 24
GB	21, 23	24, 25
PT	17, 18	19, 20

Table 4.10: Subset of sessions used to compare supervised and autonomous robot behaviour. Results comparing the performance of the autonomous robot to that of fitness instructor supervised behaviour are based on actions executed in the above 36 sessions: 2 supervised and 2 autonomous sessions per participant. Aside from DP and GB (for whom technical difficulties required a shift in experimental schedule) these sessions represent the last two supervised and first two autonomous sessions each participant undertook, to maximise similarity in e.g. session length and session difficulty.

number of each action/style across the *most comparable* sessions (i.e. consecutive programme sessions of similar difficulty) provides some insight into how well the system learned to replicate instructor behaviour with regards to utilisation of the action space. As such these comparisons were made on actions from a subset of two supervised and two autonomous sessions per participant (listed in Table 4.10). Participant evaluations of these sessions, arguably a better measure of how *good* or appropriate the robot's behaviour was, are presented in Section 4.4.3.3.

Figures 4.20 and 4.21 shows the number of each action and action style used in the selected autonomous and supervised sessions. Some clear similarities can be seen across the distributions: the *positive* style and *praise* action are those used most often and there is very similar use of the *time*, *challenge*, *humour* and *checkpre* actions, i.e. with the *humour* and *checkpre* actions being used much less than the *praise*, *time* and *challenge* actions. However there are also some differences, for example the autonomous system used the *sympathise* action more frequently, and never used the *speedup* or *speeddown* actions. This may be down to the nature of the instance-based learning employed, combined with the relatively infrequent use of these actions throughout *all* training sessions (as per Figure 4.11). However, as mentioned previously it could just be that use of these actions was not appropriate in those sessions considered. With this in mind, these results offer fairly strong support for H2A: *The autonomous robot will utilise the entirety of the co-designed action space in a similar way to the fitness instructor.*

4.4.3.2 Personalisation

Following on from the previous results is the question of whether the autonomous system was able to produce personalised behaviour across participants, similar to that demonstrated in

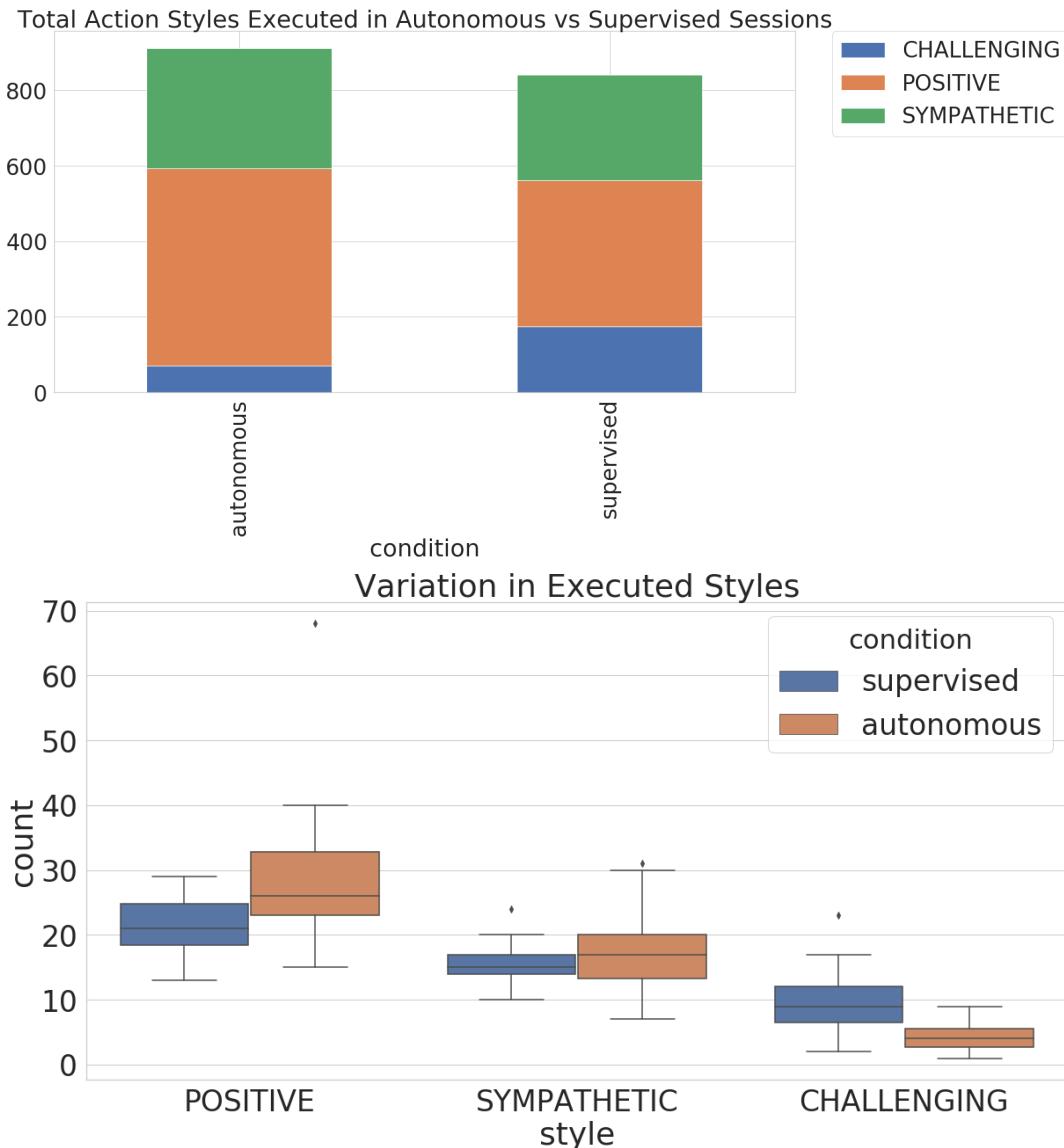


Figure 4.20: Comparison of style of actions executed in the supervised and autonomous sessions listed in Table 4.10 presented firstly as total counts across *all those 36 most comparable sessions* and secondly as a boxplot showing counts and variation across *per session*.

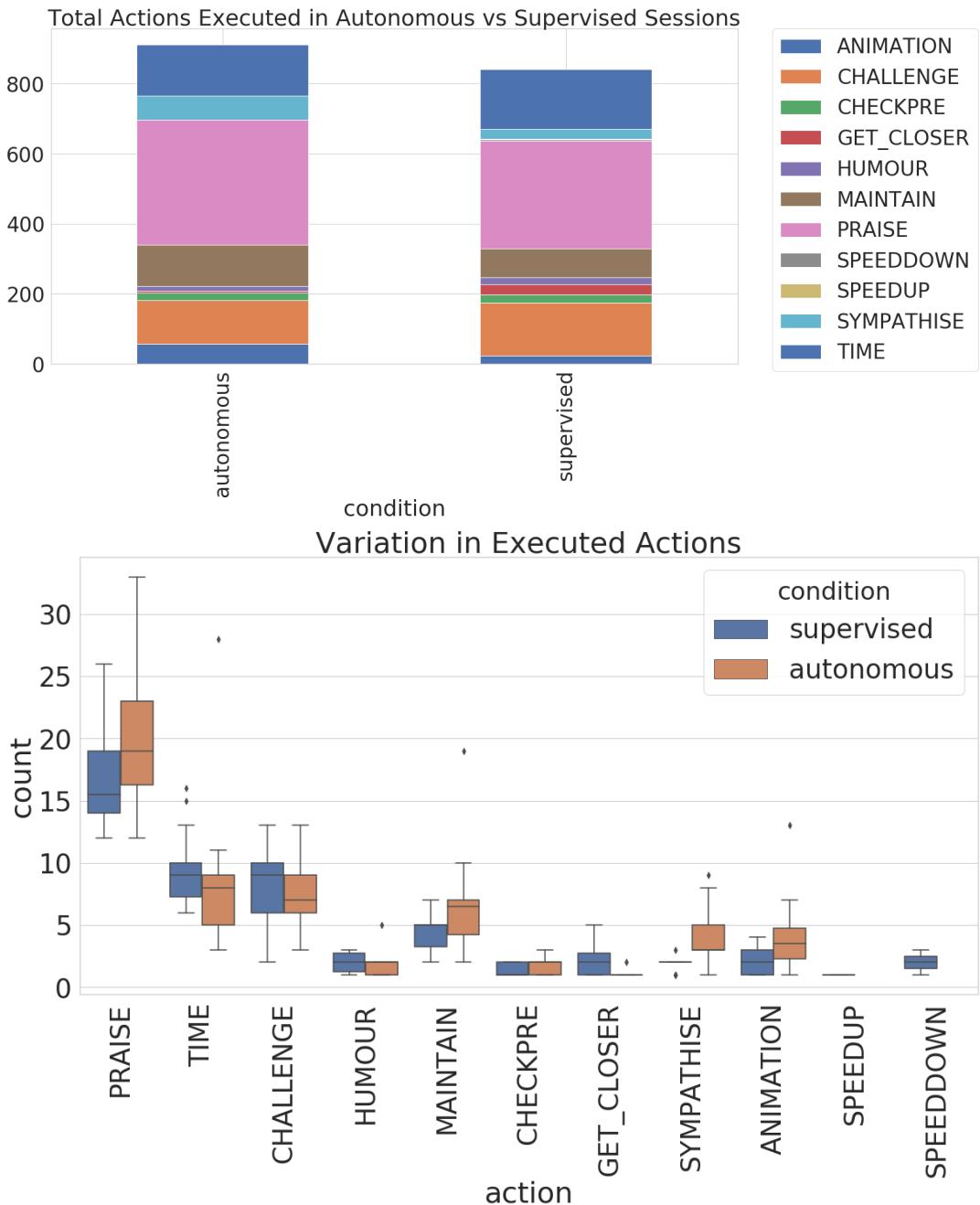


Figure 4.21: Comparison of actions executed in the supervised and autonomous sessions listed in Table 4.10 presented firstly as total counts across *all those 36 sessions* and secondly as a boxplot showing counts and variation across *per session*.

the fitness instructor's supervised use of the system. The results presented here are organised exactly as per that subsection, considering the style and action types executed across participant sessions and how those contributed to overall use of each action/style type. Again to maximise comparability, and allow comparison of raw as well as non-normalised data, actions from the two autonomous sessions per participant listed in Table 4.10 are used for making these between participant comparisons.

Figures 4.22 and 4.23 show how style and action-type of actions executed varied across individual participants' sessions. As an alternative way of demonstrating this variation, Figure 4.24 shows how full actions (combining action-type and style) executed per participant contributed to the cumulative sum of actions across all sessions. If the action/styles executed were very similar across all participants, then each participant should contribute approximately 1/9th of the cumulative sum of each action/style combination. Clearly this is not the case, e.g. with some actions not being used across all participants. Another element of personalisation that these figures fail to show is *when* which types of actions were executed. As described previously, JF and FB were two participants highlighted as requiring different approaches by the fitness instructor. Figure 4.25 shows *which* actions were used and *when* in the same C25K session (session 23) for each of them. Differences can be seen in both the overall use of certain action types as well as when they were executed. Finally, in notes taken during an autonomous testing session, the fitness instructor made reference to being impressed by the system's generation of personalised behaviour (*DP 25: 'Client specific profile is impressive.'* in Table 4.11). Together, these results provide strong support for H2B: *the autonomous robot will demonstrate personalised behaviour across participants.*

4.4.3.3 Participant Experience/Evaluation

Figure 4.26 shows participant responses to the immediate post-session question on '*How would you rate the robot as an exercise instructor based on today's session?*' firstly for the selected sessions compared in the previous subsection (as per Table 4.10) and then normalised across all autonomous and supervised sessions. Very little difference can be seen in evaluations of the robot when supervised versus running autonomously. The autonomous system received 1 'not great' rating out of a total of 32 sessions whereas the supervised system received 3 in 151 sessions.

As discussed under Section 4.3.6, Phase 3 of the testing schedule was designed in part to test whether participants would notice the (undisclosed) switch from supervised to autonomous control of the robot. Qualitative feedback collected in the immediate post-session measures of the first autonomous session suggests 2/9 participants noticed a negative change straight away:
[User FB - Session 22]: *I don't feel Pepper added much to this run. She repeated a lot of phrases and not quite at the right points.*

[User MR - Session 19]: *Usually I like Pepper's comments at the end of an intense running phase because they are usually quite short. Today Pepper said long sentences during the last 2-3 minutes*

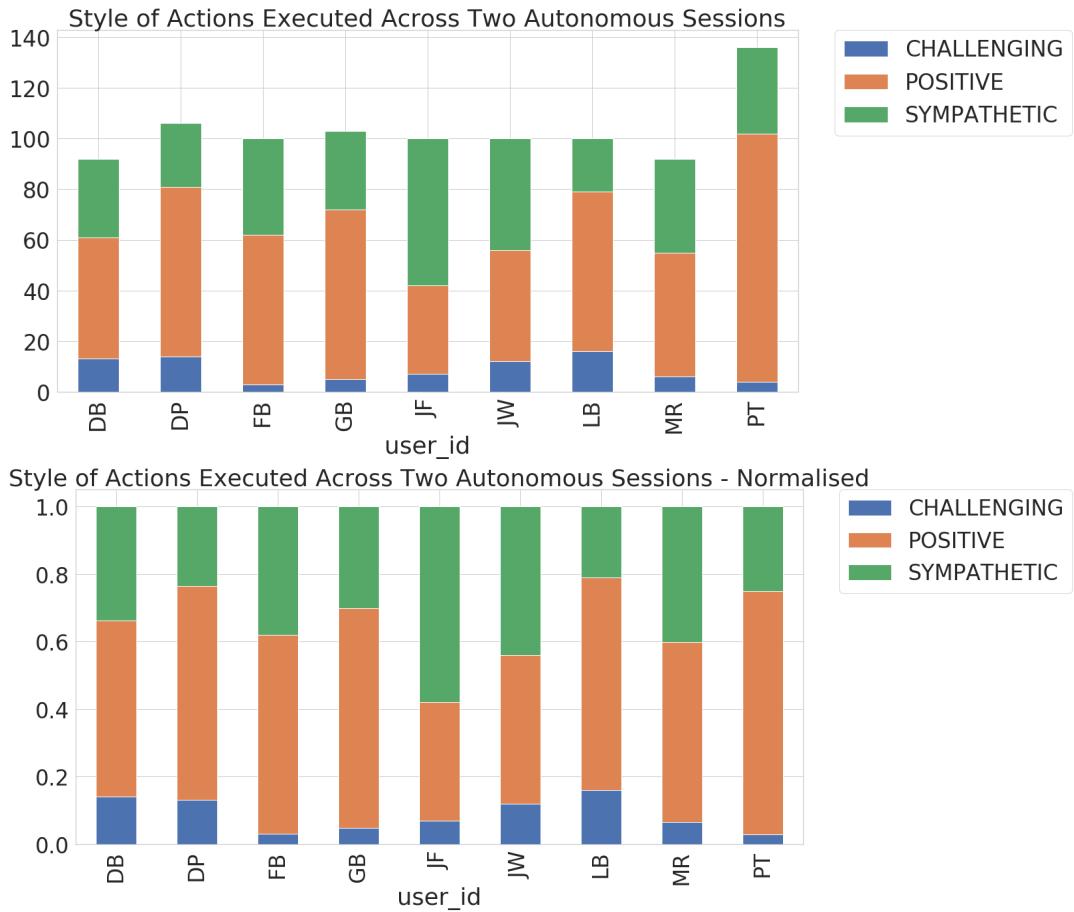


Figure 4.22: The style of executed actions across participants in the two, first (most comparable) autonomous sessions as per Table 4.10. Plotted firstly as raw data (showing the overall number of actions varied across participants) and then normalised, demonstrating the system autonomously varied the styling of actions across participants. If robot behaviour was not personalised, then we would expect the proportion of action styles to be approximately equal across participants.

of the run and that made it really hard to concentrate.

Notably, it was after this particular session that MR gave the autonomous robot the only ‘not great’ rating it received during the study (see Figure 4.26). Across all further autonomous sessions only 1 other participant (LB) explicitly referenced any change in the robot’s behaviour (negative or otherwise):

[User LB - Session 24]: *Felt a little random today- Pepper telling me to run when I was walking etc. and asking how I was doing twice in a row.*

The remaining 6/9 participants gave no indication that they noticed any change in behaviour/rated the robot any differently in the autonomous sessions. Overall, these results provide partial support for H2C: *Participants will not notice the switch from supervised to autonomous control of the robot, and will not evaluate the (autonomously running) robot significantly differently on post-session measures.* The fact that only two participants specifically identified a negative

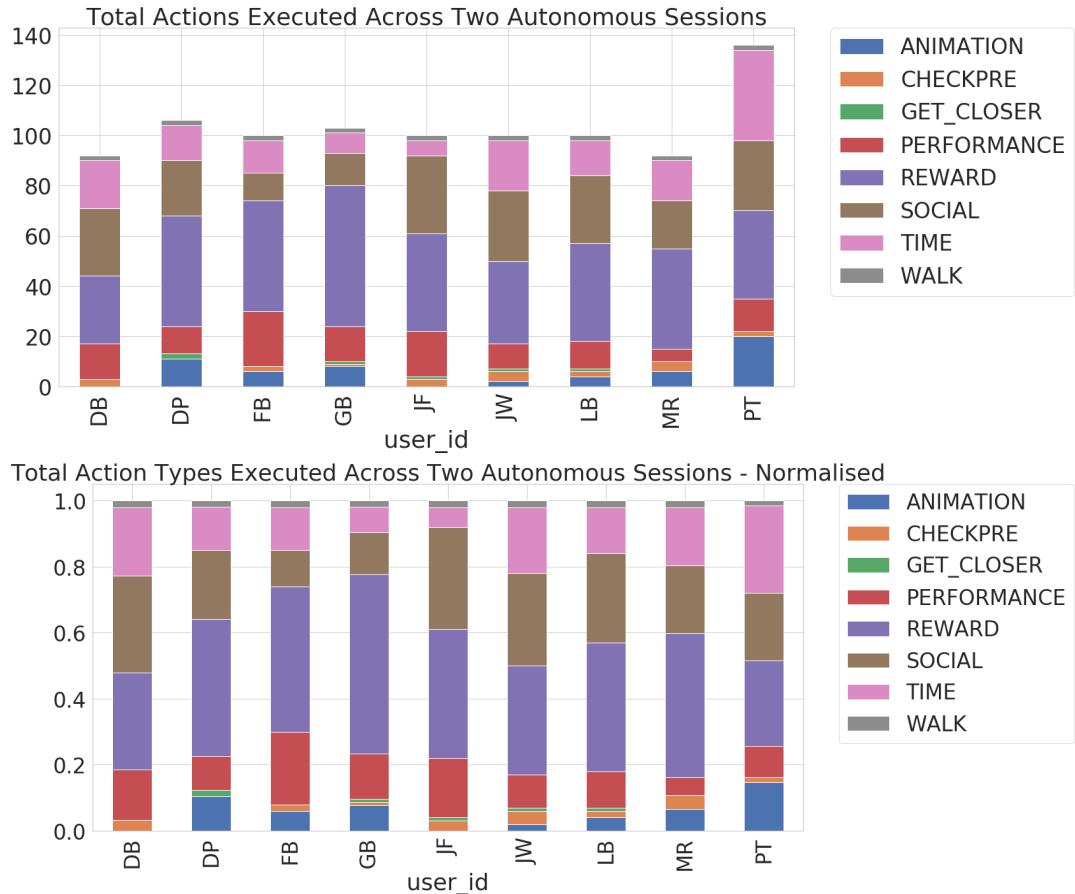


Figure 4.23: The action-type of executed actions across participants in the two, first (most comparable) autonomous sessions as per Table 4.10. Plotted firstly as raw data (showing the overall number of actions varied across participants) and then normalised, demonstrating the system autonomously varied utilisation of the action-types across participants. If robot behaviour was not personalised, then we would expect the proportion of action types to be approximately equal across participants

change gives confidence that the autonomous system was indeed somewhat indistinguishable from the supervised system. However, as this wasn't universal it raises questions as to why any differences were particularly obvious to those two participants in particular. For example, it could be that these two participants were simply more attentive to the robot's behaviour than other participants and so were always more likely to notice any changes (although this seems unlikely). More likely, it could be that the system failed to properly learn the right, *personalised* action policy for those participants in particular. This, in turn, would most likely be caused by the simple KNN algorithm employed failing to recreate the same variety of actions utilised by the instructor. This may further imply that the instructor also used less variety of actions and/or was less consistent in his use of actions for these participants, such that the respective training data were somewhat skewed.

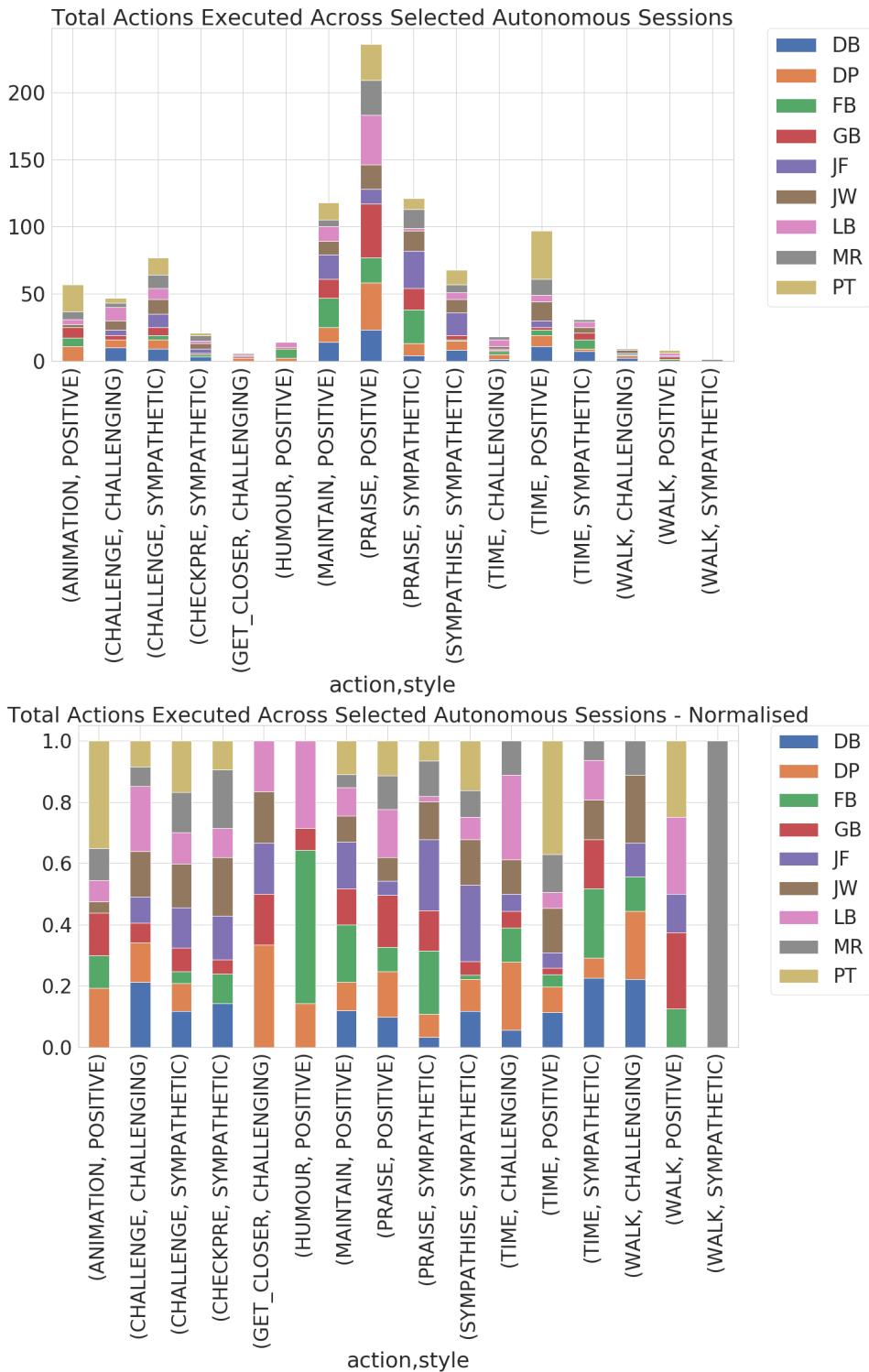


Figure 4.24: The total of each action (and associated style) executed across the two, first (most comparable) autonomous sessions participants completed, broken down by participant. The normalised version particularly shows the variation across participants; if the action distribution was the same across participants then we would expect each participant to contribute 1/9 of each bar.

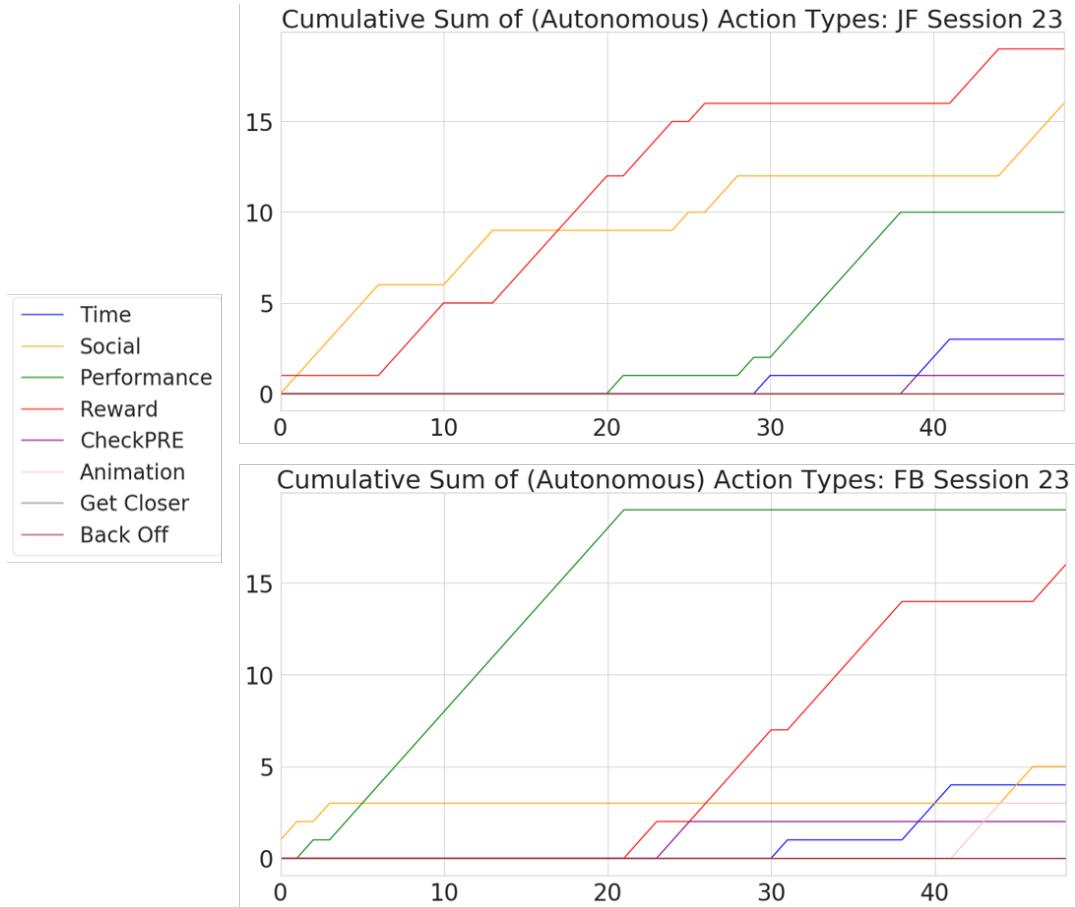


Figure 4.25: The cumulative sum of each action type throughout JF and FB's (autonomous) session 17; showing *when* each action type was used through the session to demonstrate the system's ability to personalise this element of action execution. Building on the difference in total instances of each action type (shown in Figure 4.23) it can be seen that there are also clear differences in *when* those actions were executed within the session. For example, lots of *performance* actions were used at the start of FB's session, whereas *social* and *reward* actions were used at the start of JF's session with *performance* not being used until much later on in the session.

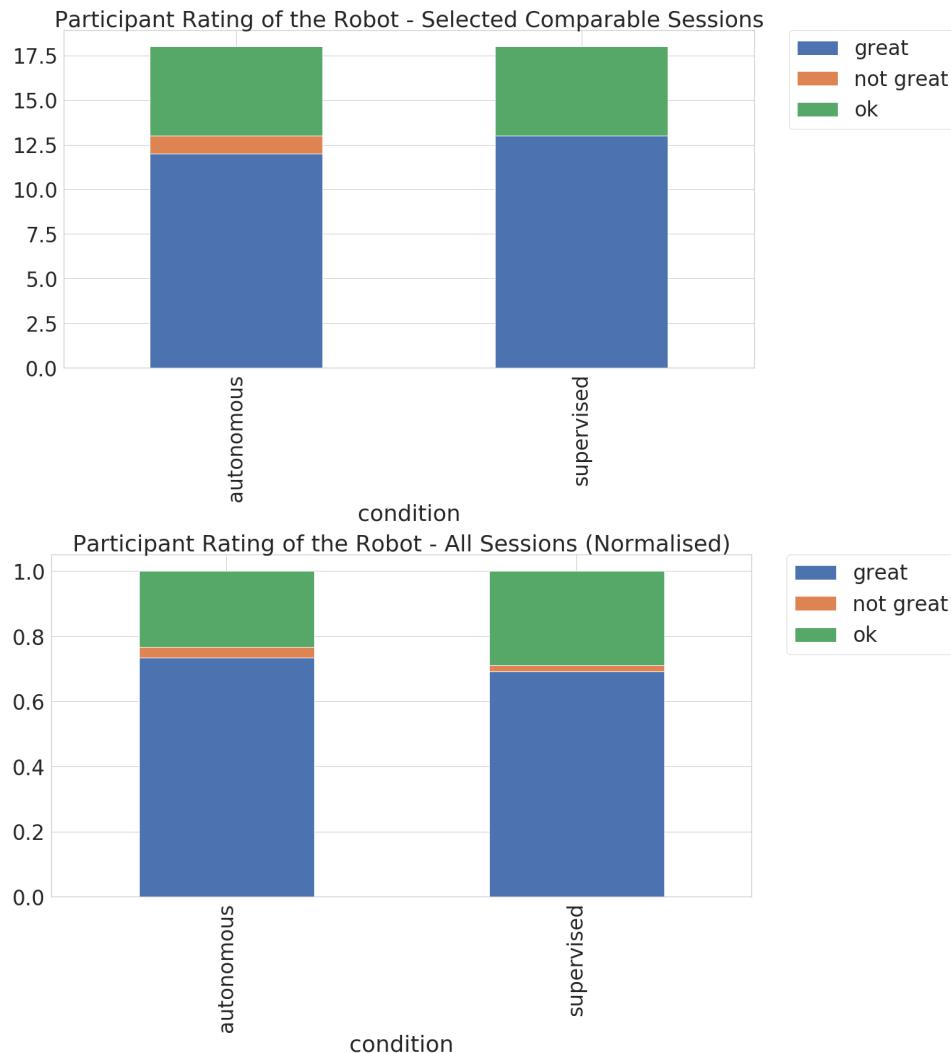


Figure 4.26: Participants' post-session ratings of the robot as a fitness instructor, first presented for the 36 sessions considered under Section 4.4.3.1 and then normalised across all (151) supervised and (32) autonomous sessions constructed during the study..

4.4.3.4 Fitness Instructor Evaluation

As noted in Section 4.3.7, it became difficult for the fitness instructor to fill in post-session measures whilst facilitating back-to-back experimental sessions. As such, his evaluation of the system is taken from notes he made in-session and an overall post-study interview. Table 4.11 gives the in-session notes, directly as written by the instructor, from the 18 autonomous sessions listed in Table 4.10. Coarsely coding these observations as rating the robot's behaviour as positive, negative or mixed, results in the distribution shown in Figure 4.27. Whilst the majority of autonomous sessions were rated positively, there were some clear instances of autonomous behaviour that the instructor felt were not appropriate. Generally these revolved around issues with repetition of actions/speech, and/or the timing of certain actions not being quite right.

In a final post-session interview, the instructor was asked to compare the performance of the Heuristic and autonomous IML behaviours. This comparison is discussed in Chapter 5, but his description of the IML generated behaviour gives further insight into his evaluation of the autonomous IML behaviour:

“With the learned A.I., it made some intelligent decisions. It like, oh, sometimes it was slow, but it made those decisions. And there are some some cases where I thought, you know, the client was really pushing themselves, they’re finding it tough. And the robot asked how they were, which was the perfect time to answer that intelligent question. And it responded to the client’s response and in the appropriate manner. It was really good... and I could tell that that level of teaching or reinforcement really, yeah really did stick, really paid off.”

In summary, considering *what* actions the robot executed *when* and for *who*, the instructor’s comments suggest the *what* and for *who* was relatively accurate, but that there were sometimes issues with the *when*. This provides partial support for H2D: *the fitness instructor will evaluate the autonomous behaviour as being appropriate and effective.*

4.4.4 Participant Experience of Couch to 5km with C25K Robot (RQ3)

Whilst this study was relatively small in the number of participants recruited, reflections on their experience still provides some insight into if/how the robot might be successful in supporting people through a long term exercise programme. Firstly, even given the typically hypothesised Hawthorne effect of participants wishing to please the researcher (Jones 1992) the fact that only 1/10 recruited participants dropped out during the study is a positive reflection on the programme and experimental setup. Low adherence to long term exercise regimes like the couch to 5km is a well documented issue (e.g. Wankel (1985), Visser et al. (2014)) and, as a reminder, participants were not reimbursed for their participation in the study in any way.

Table 4.12 presents key findings from open-answer qualitative data concerning participants’ experience as captured in the final, post-study questionnaire. Their overall assessment of working with the robot/the robot as a motivational tool ('Overall Assessment' in Table 4.12) was assessed based on their answers to these final open-answer questions. 5/9 participants specifically referred

Session	Fitness Instructor Notes
LB 23	Great starting selection of actions, choice in timings and variation. Could be more serious about challenging themselves, especially when entering the end of a program. Some sentences said after each other randomly can have contradictory meanings thus probably confusing for the client. Well timed use of humour. Good evolution of speech over the course of the run, becoming more challenging towards the end. What I've wanted to see Pepper, very happy.
LB 24	Actions: great choice (+ specific dialogue), timing and variation. Good starting combination of sympathetic challenge and maintenance (for technique), early instructions to 'zone out'. Amazing near-end-of-run speech. Much much better cooldown speech. Very happy (another) great session.
FB 22	Good variation and timing. Sympathetic time too early? Poor choice of cool down speech.
FB 23	Different start (humour focused) unique but still with good action choice and variation. Personalisation. Good smart use of speech/actions. Slightly repetitive 'maintenance' - justified? yes to be fair. Not constant, evolved throughout the run, desired learner behaviour. 2x check pre used one after the other (I'm being pedantic now). Quiet cool down = desirable. Great session.
DB 19	Happy. Praise heavy, lacking situational awareness? Sometimes with poor timing of speech revealing the robotic nature of Pepper. Not quite relevant/logical/sincere. Good versus great trainer. A lot of positive time - not enough dialogue. Better end of run speech/actions. Better cooldown dialogue/actions. Not perfect.
DB 20	Generally good action selection. Can't tell whether actions/speech is chosen at repeated intervals of time... it must be good if I can't tell? Too slow to realise the needs of the situation and react accordingly i.e. slows down for a breather/break and Pepper rewards praise/positive. Not sure there's anything Pepper could say to establish connection at this stage. Better cooldown behaviour, evolved speech/action choice i.e. becoming more challenging towards the end.
JF 23	First action too fast. A lot of sympathy and a lot of praise. Check pre with 100s left...hmm... expected better finish.
JF 24	Start on-point. Great choice and timing of actions. Slightly praise heavy..justified? Check pre better, not challenging.
MR 19	Very fast first action, some variation but praise heavy although that's definitely justified. Good positive variation, i.e. not just using or consistently using the same action especially positive time. Good timing of check pre (and in cool down). Good action choice, timing and variation. Impressive. Interesting use and timing of 'dance' [animation] action near the end of the run, don't think I've ever done that?
MR 20	Again good timing and selection of actions. Deserves praise - not too much use challenging someone already giving their everything. Therefore less need for Pepper.
DP 25	First action very fast. Good suggestions and timing. Client specific profile is impressive. Sympathetic time a little early but subjective... great dialogue.
DP 26	Great selection of starting actions. DP + Pepper: love the communication, call and response relationship.
JW 23	Very good selection of dialogue. Very happy. Choice, variation, timing = on-point. Too much check pre (x2).
JW 24	Random variation in actions? Sympathetic time far too soon. Better action choice at run end.
GB 24	Good timing and variation in actions but bad first choice of sympathetic time. Remains praise heavy, loses its sincerity? Learner keeping good variation. Amazing performance by Pepper. Good dialogue/speech choice at the end and a post run check pre. Very happy.
GB 25	Bad first choice 'not long now' with 30 minutes to go. Praise heavy but relevant. Still keeping good action variation. A lot of chat during the cooldown.
PT 19	So good at the start! Love the choice and variation... until it constantly repeats praise. Ok far too much praise. Just too much repetition. Poor finish...
PT 20	Very good start! On-point timing and choice. So impressed. Finish could be more engaging.

Table 4.11: Fitness instructor notes taken during the 18 autonomous sessions specified in Table 4.10. Coarsely coded as either positive or negative for Figure 4.27.

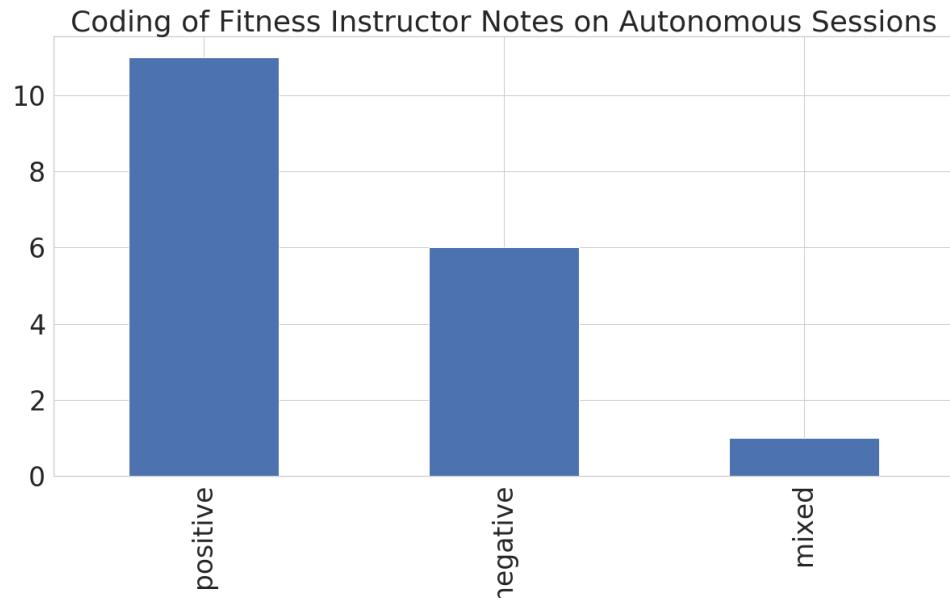


Figure 4.27: Result of fitness instructor evaluations of the autonomous sessions in Table 4.11 when coarsely coded as overall positive, negative or mixed.

to the robot as a positive motivator, whereas 3/9 questioned to some extent what difference the robot really made and 1/9 expressed an active dislike for working with the robot.

In response to the question concerning their thoughts on robot-supported exercise, 4/9 participants suggested the idea has 'potential' (but probably requires some additional development). 3/9 participants found it to be positive and useful already, based on their experiences during the study, with one participant specifically highlighting he felt confident he '*would not have completed the course if it was a mobile phone app*'. One participant suggested they were unsure how useful it would be and another was against the idea of robot-supported exercise completely.

In a measure designed to compare the Heuristic and IML systems, participants were asked whether they would work out with one of the study robots again (if so, which) or whether they would prefer not to work out with a robot at all. Only 1/9 participants indicated they would prefer not to work out with a robot in future, as seen in Figure 4.28. Together with participant ratings of the sessions, and the above described qualitative data, these results provide partial support for H3A: *overall participant experience of the programme will be positive, with specific reference to the robot as a motivational aid*. Whilst all participants ultimately expressed completion of the programme as being a positive experience, this was not always specifically linked with the robot. This variation in how useful the robot was/could be is likely down to personal preferences with regards to (i) the style of instruction/encouragement most suited to each participant and how well or not the C25K robot embodied that and (ii) to what extent the participant really benefited from any external, social motivational presence *at all* (i.e. how intrinsically motivated they were) and (iii) to what extent the participant felt comfortable with robots more generally.

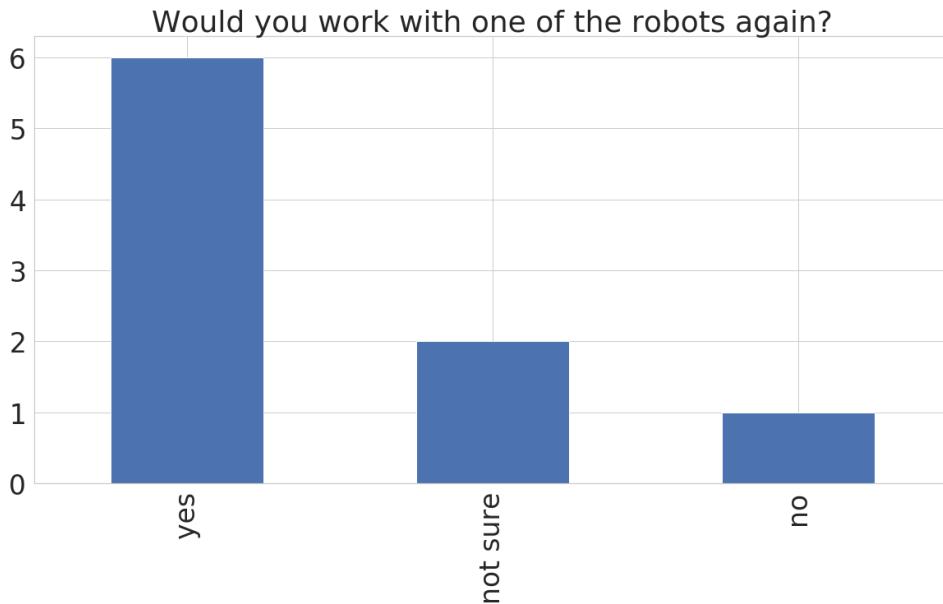


Figure 4.28: Participant responses to a post-study questionnaire item on whether they would work with one of the robots tested in the study again in the future.

Specifically addressing H3B: *the (background) presence of the fitness instructor will be a factor in participants' acceptance/positive experience of working with the robot*, 5/9 participants actively referred to the importance of the fitness instructor being present during the study. 4/5 of those participants referred to this in the context of his presence easing any safety concerns and/or giving them more confidence to work out as instructed by the robot. 3/5 referred to his presence/interactions with him as being (another) motivator in wanting to attend sessions/complete the programme. As such, there is partial support for the hypothesis. This somewhat triadic nature of interactions between the robot, instructor and participants is considered more in Chapter 5; with further detail on how participants described the relative role of the robot versus the instructor and the complimentary way in which they seemed to 'work together'.

4.5 Discussion

4.5.1 Successfully Demonstrating the Potential for SARs In-the-Wild

The experimental study designed to allow for training and evaluation of the IML system also represented a significant HRI user study in its own right. Whilst the number of participants recruited was relatively small, the study was significant in its longitudinal nature and the focus on delivery of a 'real' functional exercise programme in an ethnographically valid environment. In short, the study represented real world deployment of a SAR, ultimately allowed to run autonomously. Participant evaluations of the system, as well as more general reflections on taking part in the study essentially demonstrate that SARs may indeed be able to have a positive

**CHAPTER 4. CREATING AN AUTONOMOUS, SOCIALLY ASSISTIVE ROBOT WITH
INTERACTIVE MACHINE LEARNING**

Topic	Code	Data
Overall Assesment	Positive (5/9)	[DP] I am confident I would not have completed the course if it was a mobile phone app. [JF] I've done C25K programmes in the past but this motivated me the most to keep it up and try to do better each time. [JW] Having Pepper there... I had to push myself harder and faster than running on my own. [LB] Pepper was a good instructor and positively motivated my runs
	Neutral (3/9)	[GB] I still don't know how much effect having the physical robot there affects my motivation to run. It possibly does. [MR] I am not sure if Pepper had much influence on my performance though. [FB] I think I would have felt the same level of motivation etc. with a voice in my ear if I wore headphones whilst running.
	Negative (1/9)	[DB] I find the robot annoying over all and looked forward to just seeing and speaking with Don.
Robot Supported Exercise	Has Potential (4/9)	[MR] I feel like successful robot-supported exercise is possible. However, proper interaction between robot and human needs to be possible for that. [JF] I think it has a lot of potential for motivation if you have a robot that is intelligent and gets to know you. [JW] I do see a future in robot-supported exercise like this but I think the robot would need to learn more about the individual it was training. [GB] I think robots certainly could have a place in supporting exercise in the future.
	Positive / Effective (3/9)	[LB] I found this to be positive and could see this working well [DP] My feelings are entirely positive. I am confident I would not have completed the course if it was a mobile phone app. [PT] The robot in my opinion did a great job in helping me achieve the couch to 5k programme. It was a great gym-buddy companion that made me wanting to go to the session and try my best
	Unsure (1/9)	[FB] I'm not sure. It worked for a programme like this in terms of having additional support/motivation, however I think I would have felt the same level of motivation etc. with a voice in my ear.
	Negative (1/9)	[DB] Robot interaction isn't really for me, this study has made me realise that I need a human trainer.
Importance of Fitness Instructor (5/9)	Safety, / Confidence (4/5)	[LB] The role of Don assisted this in that having him there meant I could follow the robot's instructions safe in the knowledge that there was some support there should anything go wrong! [DP] Don's stretching routines were essential in my commitment as a major fear was dropping out because of a pulled muscle. [JW] I think the human element (i.e. Don)... was vital after sessions when aches and pains caused concern. [PT] I think I felt more secure having an experienced person... I could turn to him if I was feeling unwell during the run etc.
	Motivational (3/5)	[JF] Don's presence overall was also really motivating as he was really encouraging after each run and with the overall programme. [JW] Don and Katie were also vital in providing encouragement and incentive after each session. [PT] It helped that Don was really encouraging at the end of each session.

Table 4.12: Qualitative data collected in the final post-study questionnaire, coded using the Framework method also employed in Chapter 2 according to the hypotheses of RQ3 and emergent themes.

impact on engagement with long-term and/or monotonous exercise, but that this is only likely to be true for a subset of the population. Specifically, some people may not be affected by its presence, and some may actively be discouraged by it. Overall however, given the long-term nature of the study (resulting in high exposure to the robot, presumably going beyond novelty effects, as well as repetition of robot speech resulting from the limited dialogue programmed into the system), the results give real credibility to the concept of using SARs in this context.

4.5.2 IML for Generating Autonomous SAR Behaviour

It was stated in the introduction that a successful SAR should be able to identify *what* actions to do *when* and for *who*, and posited that IML as employed by the SPARC paradigm would be capable of achieving this. It was unknown before running the experimental study whether (i) the approach could yield successful results *at all* and more specifically (ii) whether the amount of training data collected during the supervised phases of the study (fundamentally limited by the fixed length of the couch to 5km programme and resulting experimental schedule) would be adequate.

The results presented in Section 4.4 demonstrate that the autonomous robot resulting from this application of SPARC successfully learned the *what* and for *who* but had mixed performance when it came to the *when*. The robot did appear to intelligently use certain actions at the right time with respect to, e.g. where participants were in the session and their current effort levels, however it fundamentally failed to learn a sensible action rate. Specifically, the system suggested actions for almost every input state, such that a limit for the rate of action suggestion had to be hard-coded before final testing. This is discussed further in the following subsection.

Overall however, the results point towards the approach generally being a success. The autonomous robot utilised almost all of the action space in a similar way to the fitness instructor. The *speed up* and *speed down* actions not executed by the autonomous system during testing were relatively infrequently used by the fitness instructor. This is likely because they're relatively 'high risk' actions which have the potential to negatively impact rapport between the client and the trainer and/or the client's confidence. In addition, the robot demonstrated personalised behaviour across participants. This is a particularly significant result as the previous demonstrated application of SPARC did not target any personalisation beyond responding to an individual's dynamic, in-session task performance. Finally, the autonomous behaviour of the robot was generally positively evaluated by participants and the fitness instructor.

4.5.2.1 Sometimes No Action is the *Right* Action

Given that the autonomous robot did demonstrate the intelligent timing of certain actions, its only real failure can be summarised as the inability to learn the sometimes it is best to do nothing. One difficulty in learning a sensible action rate likely comes from the fact that the fitness

instructor was also very varied in his rate of action execution, demonstrated by the example session described in Table 4.9. In addition however, it is also likely due to:

- (i) the dynamic thresholding approach utilised to apply a minimum confidence requirement to suggestions not being adequate, given the large amount of static, unchanging data within the input state space
 - and/or
- (ii) ‘*do nothing*’ not being formally encoded as an action available to the system

In both cases, these essentially reflect issues with the specific IML implementation and KNN algorithm utilised in this work, rather than the overall approach. This failure would also appear to account for the main documented criticism of the autonomous system, specifically that it was a bit too talkative/repetitive. As such, they are obvious candidates for improvement in future work, discussed further at the end of this section.

4.5.3 The Practicalities of IML as a Process for SAR Automation

Aside from the ultimate aim of generating appropriate autonomous behaviour, there are a number of practical considerations (potential benefits but also limitations) associated with this approach and demonstrated by the experimental study presented in this chapter. Firstly, the results presented in Section 4.4 demonstrate that the IML system architecture employed was adequate for enabling the fitness instructor to ensure appropriate, effective robot behaviour during supervised training sessions. In addition, participant evaluation suggests that (for those participants who found the robot to be a motivating presence) the robot was a useful motivational aid throughout the study, i.e. the robot was *effective* during the training stage as well as autonomous operation. Further, the presence of the fitness instructor acted as an additional motivator and/or reassuring presence regarding confidence in the robot for a number of participants. This latter result may be an example of inherited credibility as discussed in Chapter 3. In any case, it is another piece of evidence in favour of the robot-supported, but human-led approach to using SARs in the real world taken throughout this work.

The results also show, however, that the IML approach failed to yield any reduction in instructor workload during training sessions. Right up until the end of the fixed training period, the fitness instructor was still actively providing a lot of feedback to the system. The majority of this feedback was the refusal of suggested actions, but also included a significant amount of unprompted actions. Arguably, this is somewhat at odds with the overall result then that, when allowed to run autonomously with no such supervisory control input, the behaviour was actually appropriate and effective (and evaluated as such by the instructor himself). This observation results in two key considerations for the practical application of this approach.

Firstly, it raises doubt on the possibility that the SPARC approach might *usefully* reduce the workload of the person teaching it and/or supervising it when running ‘autonomously’

if that teacher is still to continue supervision for e.g. safety reasons. In this work, during supervised sessions, the fitness instructor's attention was generally completely taken up by monitoring participants' interaction with the robot and monitoring of the teaching interface. Further, this was required not just for the generation of unprompted actions but also for the *arguably* 'lower' workload of supervising the system and responding to inappropriate actions. For a fitness coach, the appropriateness of the action depends on a complex multitude of factors regarding instantaneous but also preceding performance and interactions, thus assessing that appropriateness requires significant involvement and attention from the instructor. However, this may not be true for other applications in which (i) inappropriate or unsafe actions might be easier to spot by someone supervising the system and (ii) the supervision interface is designed around alternative modalities such that visual attention of the supervisor is not required at all times.

Secondly, it raises doubt on the ability of the expert supervisor to actually assess at what point the learning system is '*good enough*' to be allowed to operate autonomously. The previous demonstration of SPARC also demonstrated this same result of the system yielding no significant workload reduction yet generating appropriate autonomous behaviour (Senft et al. 2019). In this work, the amount of training time was fundamentally limited by the experimental schedule. However it is completely feasible that other applications of this approach would leave it to the expert supervisor to decide at what point the system requires no further training data. Results from two studies can not be used to suggest an overall trend, but it is interesting to consider the notion that whilst supervisors *can* control/adjust the robot's action policy then they *will*. Future work might investigate this further, and if it is the case, consider how it might be accounted for within system design and implementation. To really comment on the effectiveness of the learning process, future work might also consider a comparison to the equivalent expert led teaching of a new (*naive*) *human* instructor.

4.6 Conclusion

This chapter presents the design, implementation and evaluation of SAR, automated via interactive machine learning carried out in-the-wild. The system architecture employed is technically novel, building on previous work by (i) including *static* participant personality/motivation data in the system input space and (ii) utilising a dual-learner approach that allowed for generation of a robot *style* or '*mood*'. The *style learner* was then used for informing lower level robot behaviours as well as styling actions, with both of these extensions being critical in the learning of personalised action policies.

The in-the-wild study utilised for training and evaluation of the system represents a significant HRI user study, both based on its longitudinal nature with many repeated participant-robot interactions but also in its ethnographic validity and functional delivery of a real world useful

programme. Key findings from the work can be summarised as follows:

- The success of the overall user study setup with regards the role of the robot and the instructor and his use of the IML system demonstrates the value of the co-design process employed early on in the work.
- The importance of personalisation in the context of social assistance, already highlighted by the studies in Chapters 2 and 3, was demonstrated in practice, and successfully achieved by the autonomous system.
- Further regarding personalisation, and in-line with assumptions/limitations identified at the beginning of this chapter, it is clear that a SAR may not be acceptable or appropriate for all potential users, simply down to their personal preferences regarding robots. In addition, some users may benefit from a SAR that takes an *alternative social role* e.g. less authoritative to the one considered here.
- The presence and role of the fitness instructor proved to have an important impact on overall participant experience, including their interactions/willingness to work with the robot. In contrast with traditional HRI studies that look to minimise any additional human/social presence, it is argued that this is both crucial but also methodologically valid for considering robots that will be deployed in the real world.

Overall, the results of this user study give strength to the idea that SARs might positively influence people to stay engaged with a long-term, monotonous exercise programme. In addition, this work demonstrates that IML offers a feasible method for generating complex, personalised, autonomous SAR behaviour using input from a (non-robotocist) domain expert. However, future works should carefully consider how the concept of *doing nothing* might be encoded and/or learned in an IML system. Further observations on IML as a design process, and its suitability for pursuit of a mutual shaping approach to SAR design and evaluation, are presented in Chapter 5.

MUTUAL SHAPING IN DESIGN AND DEPLOYMENT OF SOCIALLY ASSISTIVE ROBOTS

As discussed in Chapter 1, an aim of this work was to undertake a *mutual shaping* approach to the research and development of socially assistive robots, positing that such an approach is necessary for the resultant robots to be effective when deployed in the *real world*. This chapter identifies (i) how the (generalisable) methodologies employed in this work support mutual shaping, (ii) why such methodologies are worthwhile and (iii) examples of mutual shaping ‘in action’ taken from the work presented across Chapters 2 to 4. Part of the work presented in this chapter (specifically limited to the focus group methodology and study with therapists) is described in the following publication:

Winkle, Katie, et al. "Mutual shaping in the design of socially assistive robots: A case study on social robots for therapy." International Journal of Social Robotics (2019): 1-20.

5.1 Introduction

Chapter 1 introduced the concept of mutual shaping, which can be summarised as the two way interaction between a robot (or more specifically, use of that robot) and the broader social context or environment in to which that robot is deployed. Further, it was identified that this work would take a *mutual shaping approach* throughout, specifically by employing participatory methods where possible and considering mutual shaping effects regarding deployment of socially assistive robots (SARs) in the real world (Sabanovic 2010).

The focus group methodology utilised in Chapter 2 was designed specifically to support this approach. The novel methodology employed resulted in a number of additional observations, specifically concerning (i) the potential for mutual shaping effects on deployment and (ii) evidence

of mutual shaping achieved *during/as a result of* the study itself. These results are presented and discussed in Section 5.2. Similarly, one reason for utilising the interactive machine learning (IML) methodology presented in Chapter 4 was to go one step further than traditional participatory/co-design and allow *direct* user/stakeholder participation in the *automation* of the robot. As briefly introduced in Chapter 4, a heuristic, rule-based robot was also co-designed and tested alongside the IML system during that study, in order to investigate the differences in robot behaviour (and participant experience) resulting from the two approaches. This is discussed in detail under Section 5.3.

This chapter contributes two methodologies for *how* to take a mutual shaping approach during SAR design and development, along with results that demonstrate *why* that's a good thing to do. In addition, observations of mutual shaping collected during this work are also presented as evidence of the complex interactions between robots (and robotics research) and the context of use. From the study with therapists of Chapter 2 this represents the extended focus group methodology (*how*), the insightful observations it yielded (*why*) and the actual impact it had on participants' acceptance of robotic technologies (example of mutual shaping). For the C25K robot coach system presented in Chapter 4, this represents interactive machine learning as a participatory design process (*how*), the results demonstrating that the resulting system was better than an expert-informed heuristic based system (*why*) and observations regarding real world deployment and use of this system over a longitudinal study (examples of mutual shaping).

5.1.1 Methods for Mutual Shaping

There are a number of methodologies that might be employed in a mutual shaping approach to SAR design and research, and it is useful to define them in order to properly situate this work amongst existing literature. These include:

1. *Ethnographic / 'In-the-Wild' Studies* typically focus on understanding situated use and/or emergent behaviour(s) on deployment of a robot into the real world. Concerning robot design, such studies are inherently limited to the testing of prototypes. However, they might be used to inform initial design requirements through observation of the current use case environment and user behaviour.
2. *User-Centered Design* aims to understand and incorporate user perspective and needs into robot design. Typically researchers set the research agenda based on prior assumptions regarding the context of use and proposed SAR application.
3. *Participatory Design* encourages participants (users, stakeholders etc.) to actively join in decision making processes which shape robot design and/or the direction of research. This typically involves participants having equal authority as the researchers and designers, with both engaging in a two-way exchange of knowledge and ideas.

These terms define relatively high level methodologies or research philosophies. Specific data gathering methods which can be employed in pursuit of the above include focus groups and workshops (e.g. Jenkins & Draper (2015), Louie et al. (2014), Lee et al. (2017)), interviews (Lee et al. 2017), surveys (Green et al. 2000) and observations (e.g. Sabelli et al. (2011), Forlizzi et al. (2004), Chang & Šabanović (2015)). Lee et al. (2017) give a good overview of the above practices as currently employed in robot design, with a focus on user participation in the design process. On the distinction between user-centered design and participatory design, they note that user-centered design typically tries to understand user needs for informing robot design. Participatory design instead attempts to empower participants such that they can actively collaborate in the design process. The authors also describe the concept of mutual learning, a key mechanism for achieving such collaboration. Via mutual learning, users learn about design and technology from the researchers as well as providing useful information and perspective to the researchers. This empowers the users to be able to really take part in the design process, giving them the knowledge required to conceptualise and hence critically evaluate the concepts proposed. Following these definitions, the focus group methodology presented in Section 5.2 utilises elements of both user-centered and participatory design, with a focus on mutual learning. Section 5.3 then discusses how expert-in-the-loop, interactive machine learning can be considered a participatory design process, and also utilises significant ‘in-the-wild’ robot deployment.

5.1.2 Related Work

Most studies concerning the development of SARs for exercise engagement have been concerned with feasibility and quantifiable impact (e.g. Gockley & Mataric (2006), Tapus & Mataric (2008)) rather than exploring use cases, generating design recommendations or considering mutual shaping effects. Studies designed to measure user acceptance have also typically followed a technologically deterministic approach, with a complete system being presented for evaluation. For example, in the a closely related work considering SARs specifically for rehabilitative exercise engagement, Wilk & Johnson (2014) utilised a robot demonstration in investigating the potential for a combined telepresence/SAR system in facilitating and encouraging engagement with stroke therapy. Residents and caregivers from a daycare centre were given a demonstration of the robot’s capabilities. Then, they were asked to complete a survey measuring perception and acceptability of the robot system. The authors note that caregivers also discussed additional capabilities the robot could have, but no detail is given as to the format or formality of these discussions. Further, there wasn’t any consideration of, nor any opportunity to explore, mutual shaping effects that might arise through deployment of the system. Similarly, whilst caregivers identified potential robot applications, how the robot would be incorporated into overall care delivery was not discussed.

Considering SARs more generally, other research considering robots for the care of older adults has typically employed user-centered design to elicit user views or assess user needs

for informing design requirements (e.g. Louie et al. (2014), Wu et al. (2012), Beer et al. (2012)). Whilst these works provide valuable user insight, they do not amount to pursuit of a mutual shaping approach because they fail either to i) account for the influence of social context on robot deployment and/ or ii) allow societal influence on robot design or research direction during the development stage. Other works however have specifically employed a mutual shaping approach, either through observing users and robots in real social environments via ethnographic studies (Sabelli et al. 2011, Forlizzi et al. 2004, Chang & Šabanović 2015) or through attempting to actively involve users in robot design (Lee et al. 2017, Azenkot et al. 2016) or the shaping of robotics research (Jenkins & Draper 2015).

Specifically using participatory design, Azenkot et al. (2016) generated design specifications for a SAR that could guide blind people through a building. The authors' study consisted of multiple sessions including interviews, a group workshop and individual user-robot sessions. Initial interviews were used to brief participants about the research and robot capabilities. The group session was used to develop a conceptual storyboard of robot use, identifying interactions between the robot guide and the user. Finally, participants were individually invited to work with a researcher and robot platform to prototype robot behaviour. The researchers also asked participants to instruct a naive human guide, asking probing questions around their preferences and instructions as a form of contextual inquiry attempting to elicit tacit knowledge.

Somewhere between user-centered and participatory design is the Jenkins & Draper (2015) use of focus groups to explore views on care robots. The authors used focus groups to collect views on ethical issues stemming from the real world deployment of care robots from older people and their carers. Participants were presented with pre-designed scenarios in order to prompt discussion, which may have limited the scope of discussions but ensured that the context of use was well established. This meant that the use of such robots was considered very holistically in terms of real world situations. In addition, participant responses were used to prompt discussion on how the SAR-integrated care approaches could be adjusted in order to make them more acceptable and reduce negative consequences on deployment. This work (and the larger research project from which it stems) fits somewhat with the mutual shaping approach in its attempt to consider SARs in care holistically and in allowing stakeholders to shape their work through revision of the pre-designed scenarios between iterations of the focus groups. Similar considerations about how therapy is currently delivered, and how that may change, were made in the study with therapists of Chapter 2, and are discussed under Section 5.2.

Particularly relevant to our work is the way in which Lee et al. (2017) used participatory design methods in the development of SARs for older adults with depression. The authors present a multi-stage participatory design process including interviews and workshops designed to facilitate co-design. They note that the use of multiple sessions allowed participants and researchers to 'familiarise themselves with each other's knowledge and build a relationship of trust'. Initial interviews were used to give researchers an understanding of the context of

potential use (i.e. by visiting older adults in their home to see housing arrangements) and to encourage initial development of trust between the participants and the researchers. These were followed by a number of workshops with the older adults. The first workshop was used to introduce participants to examples of robot systems and to explore how they might be used. The second and third workshop focused on generating robot designs, after asking participants to reflect on some element of their life (challenges faced when lonely and use of space in the home respectively). The fourth workshop focused more on technical design with participants suggesting sensors that could be included in a robot and discussing resulting issues around data collection, sharing and security. Finally, a fifth workshop was used to share the robot designs generated by the older adults with therapists to get their perspective on such robots being used in the older adults homes. Elements of this work are reflected in both the focus group methodology and interactive machine learning process described below. The focus group methodology shares Lee et al.'s focus on mutual learning and utilises the same approach of introducing/explaining robot systems and getting participants to reflect generally on their current situation before encouraging participants to translate that into robot design requirements. However, the aim was to achieve this in a single focus group session. The interactive machine learning methodology shares the focus on co-design across multiple design sessions based on a close working relationship between the researcher and an expert stakeholder.

The above literature demonstrates that whilst there is significant effort to include users in robot design, this is often achieved using user-centered design methods. Such methods typically focus on eliciting user perspectives in a one way exchange. Users provide information which designers and researchers then incorporate into robot design (e.g. Louie et al. (2014), Wu et al. (2012), Beer et al. (2012)). Wu et al. (2012) note that their focus groups, designed to identify design preferences for robots for older adults, 'offered an opportunity for participants to...challenge some implicit preconceptions of the roboticists'. One clear way the work in this chapter builds on this is to additionally consider exactly the reverse; how roboticists can challenge participants preconceptions, attitudes towards and acceptance of SARs through a research study. None of the aforementioned studies do this, although Lee et al. (2017) do reflect on the potential participant benefits of participatory design (e.g. empowerment, increased social interaction) and how this should be considered in its use. Even those studies specifically following a mutual shaping framework have typically been focused on the use of ethnographic studies to generate observation data and understand mutual shaping of robot use on deployment (Sabelli et al. 2011, Chang & Šabanović 2015) rather than during the design process, and hence gives little opportunity for users or other stakeholders to voice their perspectives during SAR design. Works employing participatory design (e.g. Lee et al. (2017), Azenkot et al. (2016)) are the exception, and demonstrate the worth of using such approaches.

This literature demonstrates to what extent engagement with stakeholders has been limited to either (i) informing robot design guidelines to feed into researcher/engineer-led development

and/or (ii) consideration of ‘finished product’ systems. This chapter novelly describes how interactive machine learning can be employed as participatory design to address this, offering a way for domain expert(s) to actively contribute to the *automation* of robot behaviour.

5.2 Extended Focus Groups for Mutual Shaping

As discussed under Section 5.1.2, previous work in HRI have used focus groups to engage stakeholders in user-centered robot design (e.g. Louie et al. (2014), Wu et al. (2012), Beer et al. (2012)). However, such focus group studies typically represent a one-way exchange with researchers simply looking to extract knowledge, information or ideas from the participants. The methodology described here, and utilised in the study with therapists presented in Chapter 2, is instead extended toward more participatory design through use of mutual learning.

5.2.1 How: Extended Focus Group Methodology

The key elements required for focus groups that support mutual learning between the participants and researcher(s), alongside examples of research questions/prompts from the study with therapists presented in Chapter 2, are given below. Table 5.1 further identifies how these elements were reflected in the overall structure and topic guide of the focus groups undertaken during that study. Firstly, taking part in such focus groups has a significant impact on participants themselves (likely due to the mutual learning element and focus on their inclusion in the research). Secondly, it allows for broader issues concerning the proposed robot application to be raised and considered early in the design stage. Finally, whilst not demonstrated in this work, such focus groups could also be used to as a tool to re-consider these issues once the robot (or e.g. a prototype) has been deployed for testing in the real world. Overall, the methodology aims to:

- Establish participants as experts and engage in broad discussion without presenting a defined research agenda (as per participatory design methods).
- Get participants to reflect on the context of use as it is now; i.e. before introduction of a robot, in order to ground discussions.
- Take time (in the middle of the session to first allow for item 1) to explain the research agenda (and motive) in more detail, as well as to share relevant technical expertise and showcase robot capabilities to improve participant understanding of what’s possible.
- Revisit key topics of discussion after the above in order to get informed (and/or altered) opinions, targeted around the research question(s) and proposed application(s), as well as any new ideas (as per user-centered design methods).

5.2. EXTENDED FOCUS GROUPS FOR MUTUAL SHAPING

Section	Topic Guide	Key Features
Discussion Part 1	Conventional therapy delivery Use of social robots in therapy Activity on factors affecting adherence	Expert establishment Naive unconstrained discussion ML1: Domain expert-to-researcher
Research Presentation and/or Demonstration(s)	Study motivations & related literature Project aims/objectives 2x Pepper demonstrations	ML2: Researcher-to-domain expert
Discussion Part 2	Use of social robots in therapy	Informed, targeted discussion

Table 5.1: Key elements of focus group methodology as applied to the study with therapists. ML = Mutual Learning.

5.2.1.1 Key Elements for Focus Groups which Support Mutual Learning

The key elements required for the methodology, as referenced in Table 5.1, are explained below. A related topic guide item/focus group activity from the study with therapists in Chapter 2 is given with each to demonstrate how they can be implemented.

1. Expert Establishment

A small amount of time at the beginning of discussions should be dedicated to making participants feel comfortable. Specifically, the researcher should raise a question or topic of discussion that all participants will feel well-qualified to answer (with minimal hesitation/reluctance to make suggestions). Qualitative research guidelines (Curry 2015) suggest that expert establishment in focus groups is key to encouraging participation. Participants then feel confident that they can offer useful, valid contributions and are therefore less hesitant to take part. Example: *What are your main goals when working with service users?*

2. Naive, Unconstrained Discussion on Robots for Proposed Application

This part of the discussion allows solicitation of participants initial ideas on the use of robots for the proposed application, before they can be biased by the researcher's somewhat pre-defined research agenda. In order to ground discussions, it might be helpful to provide a brief description/images of the proposed or example robots, but this shouldn't include anything about the proposed use/application. Example: *(Participants were first provided with a collage of images and a definition of the term 'social robot') What do you think about using robots to support a therapy program? How do you think that robots might be able to do that?*

3. Mutual Learning 1: Domain Experts -> Researchers

This part of the discussion is focused on allowing the researcher to get an informed insight in to the realities of the application domain. This can be somewhat targeted to factors that might most impact/inform robot design and functionality for their proposed application.

Example: *How do you monitor service users' engagement? What kind of factors do you think affect service users' compliance with self-practice exercises?*

4. Mutual Learning 2: Researchers -> Domain Experts

This part of the discussion is essentially the opposite of the previous, and is focused on the researcher sharing their proposed application, as well as their reasoning and rationale behind that, with the participants. In addition, it can be used to demonstrate and explain robot capabilities to further inform participants such they have a better understanding of what such a robot might be able to do. Example: *Researcher-led presentation of background literature demonstrating positive impact of SARs on exercise engagement, overview of larger doctoral research aims, objectives and planned activities and 2x robot demonstrations depicting possible application behaviours.*

5. Informed Discussion on, & Revisit of, Robots for Proposed Application

The final part of the discussion allows for a re-visit of the initial topics for more informed and more targeted discussions (i) regarding the researcher's proposed application and/or (ii) based on participants increased understanding/familiarity with robot capabilities. Example: *What do you think of the demos? How would you have done it differently? Now that you have seen the demonstrations... what would a robot aid look like, how could it help?*

5.2.2 Why: Insightful Results Concerning real world Deployment

Chapter 2 demonstrated how results from these mutual shaping focus groups fed in to the development of design guidelines for SARs in therapy. However, a number of additional observations concerning real world deployment were also collected as a result of (i) the initial, un-biased discussions and (ii) the focus on mutual learning specific to this focus group methodology. These results are presented here to therefore demonstrate the worth of this approach.

5.2.2.1 SARs in Therapy: Mutual Shaping Issues Identified for Consideration

Across all focus groups participants identified a number of social and societal factors relevant to conventional therapy delivery. These particularly included factors which influence patient engagement and how the therapists worked to address that. Such factors are also likely to affect SARs deployed in therapy, and in some cases participants made this inference themselves. One key recurrent theme was the influence of (and potential to affect) the patient's immediate social circle, who are often a key support to both the patient and the therapist. A second key theme was the importance of the therapist-patient relationship and communication, both in terms of how this is crucial to conventional therapy delivery but also how SAR use might facilitate or be influenced by it. Another key theme was patient demographics and costing; although detailed financing was generally avoided by the moderator, this raised some pertinent issues around patient socio-economic status with regard to therapy engagement. Finally, the potential for

patients to become dependent on the robot was raised across 4 out of the 5 focus groups. These themes are discussed in more detail below.

Issues Regarding the Patient's Social Life, Circle & Support

The role of immediate family in encouraging the patient to engage was acknowledged multiple times:

[P2]: *"His wife will always come back to me and say what he has and hasn't been able to do...she likes to make sure he's doing everything he's meant to do and sometimes he's sitting there and he's like 'well I tried to do it'"*

Some warned however that, particularly if the therapy was initially instigated or particularly encouraged by a particular family member, too much of this co-operation between the therapist and family could isolate the patient:

[SL1]: *"if the person around them is the one that's referred them and you're seen as that you're coercing, it's a bit of a conspiracy then isn't it and the person's going to feel a bit left out"*

Some participants also noted the strain that therapy can put on personal relationships, and similarly whilst social support could be key in encouraging engagement, it could also have a negative impact as well:

[P2]: *"he's a good example of someone I've seen and given him exercises to do and challenges to do when I'm not there...he's got someone with him there who enables him to do it but he sometimes gets a little bit irritated because it's his wife (laughs)...And that's not uncommon either"*

[P5]: *"sometimes the patient won't want to do the exercises with their for example husband because they know he's already doing all of the activities that they used to do and they don't want to be that additional burden: 'they've already had a busy day I can't ask them to do my exercises with me so I just can't do my exercises'"*

Participants were enthusiastic about the use of SARs to somehow address this, considering how SARs for therapy in the home could reduce carer load and relationship strain, but also raised potential concerns, e.g. the potential for guilt, that might be associated with that:

[SL3]: *"Parents, carers, they are absolutely knackered...we spend our lives telling them not to sit their children down in front of the telly to look after them so I think there might be, there's an issue of this around that as well like 'ooh I'm handing my child over to a machine to do what I should be doing'...I think it could help but there could be a guilt loop as well"*

Such discussions generally focused on the SAR becoming a 'third-party'; i.e. something which could neutrally prompt the patient or alternatively be more convincing than a family member (especially in the case of young people):

[OT4]: *"Just thinking about I suppose the child-parent dynamic...actually the parent sometimes, you know, children say they're quite tired after school and parents like 'oh you don't need to do it tonight' or the opposite they try to get them to do it and they're like 'I'm not going to do that'. So actually with kids I think it could be a useful tool actually...a little bit more independent"*

CHAPTER 5. MUTUAL SHAPING IN DESIGN AND DEPLOYMENT OF SOCIALLY ASSISTIVE ROBOTS

[SL3]: “*But then also a lot of people respond to a professional and not their family... So you could have the spouse sort of whistles to the robot ‘Go on, give him a shout’*”

How We Would, Could or Should Talk to Robots

The concept of whether or not manners should be required or understood and responded to by a SAR was brought up both in terms of how the user would want to speak, but also about how that could affect or influence people around the robot:

[SL1]: “*I think bits of me like the human aspects but bits of me want it to almost to be a computer that talks*” (agreement from others) “*because I’m not sure... I would want to say please, and then I would hate myself for saying please to a machine*”

[OT7]: “*I don’t like speaking to Siri because it doesn’t recognise when I say please, it then confuses it... when I’m talking in the car and my kids are hearing it I don’t want them hearing it as if it’s an instruction*”

Therapist-Patient Relationship & Communication

All participants noted the importance of the therapist-patient relationship, e.g. describing the importance of rapport and the influence that therapists can have on patients:

[P2]: “*as therapists we’ve got the, well, sort of luxury of having time with people, so you do build up a relationship and mostly there’s quite a lot of trust there and they put quite a lot of trust in what you say so you can be very influential with it*”

A suggested benefit of the system was addressing individual patient preferences with regard to monitoring and disclosure. For example, patients who didn’t like reporting to the therapist could instead report to the robot, whereas those who seemed to benefit from therapist reinforcement could be reminded of that by the robot:

[P5]: “*For some patients the idea that it’s not a human that you’re reporting to, and it would be faceless entity, could be a benefit to them, and them knowing that someone else is going to read it and observe it could be an issue for them... I would say it does make a difference because some of the technology that I use where it says about reporting the number of steps, they’re waiting to be told off, even though you don’t tell them off, or waiting to be praised.*”

One participant also noted that patients might be more willing to ask questions or ask for help via the robot based on previous experience with an email based system:

[OT6]: “*...what it also does is enable the patients to respond and write back in to their therapist with questions ‘should I, shouldn’t I, how do I do it’ and in fact we have had a couple of patients who’ve been very engaged with it and even do so from their hospital bed and that was really interesting in terms of opening up and at what stage they share information and do those things and because there wasn’t somebody there in front of them, which wasn’t as off putting then actually you got very different information.*”

Robot Dependency

Four participants independently referenced the risk of patients becoming dependent on the robot, having less autonomy or feeling less responsibility for managing their condition:

[OT3]: “*if she can go and get him a cup of tea then she would be loved.*” [OT2]: “*That would be great*” [OT3]: “*but at the same time your patients aren’t getting mobile they’re not getting up and aren’t getting up and engaging.*”

[P6]: “*it’s really important to keep the responsibility with the person themselves, which probably is going to be a factor with the robot as well you want to potentially be able to remove the robot from the situation at some stage*”

Cost & Patient Demographics

Given that cost is a typical barrier to adoption of new technologies, it was decided that issues relating to these aspects would not be discussed in depth. However, some participants highlighted more specific issues regarding cost. In the first instance, in the groups containing private practitioners, there were comments suggesting that patients who paid for their therapy directly were more likely to engage with their programme:

[SL3]: “*I find that people do the work more in private practice than they ever did in the NHS*”
[OT6]: “*But then, the fact that they are paying means that they have a vested interest*”

In addition however, when considering factors which affect engagement (in which patient demographics was highlighted as a factor), some participants identified that those who might struggle with motivation most are least likely to be able to afford private therapy or related technologies:

[SR2]: “*I was at a sports medicine conference...and [a sports science professor] was talking about how actually interventions that we use to try and improve people’s health through basically trying to improve their motivation to exercise...he’s saying we’re trying to solve sort of lower class problems with middle class, upper class solutions with things like apps and things you know...these are people who potentially don’t have smart phones and yet all our efforts are being pointed at things like that technology which can’t be afforded to these people*”

Willingness to Work With/Adapt to Using SARs

During the pre-demonstration discussions, all participants indicated they would essentially be willing to try anything which might have a benefit for patients.

[OT6]: “*I think if you work with people, patients then anything that makes a positive difference preferably (laughs) is worth considering*”

One participant offered an interesting reflection concerning what’s best for therapists versus what’s best for patients, and how those two things might unintentionally be misaligned, based on a previous experience working with technology in therapy:

[SL2]: “*I did a project quite some years ago about using video conferencing for face to face sessions and what was really interesting was that therapists were dead against it, including me, and felt that that would erode our one to one thing... and we really weren’t very keen, people really*

weren't very keen, myself included. The patients loved it, so suddenly certainly for me I had to make a real leap into 'ok...I thought I was thinking of my patients' best interests but what I was actually thinking was about my satisfaction and the rewards for me' so I really had to change that and...so I'm probably coming at this much more positively than I perhaps would have done, having had that experience."

One participant raised the issue that in the case of patients experiencing pain, the robot would have to tell them to stop, whereas a therapist would assess the situation and (if appropriate) reassure them that some pain is to be expected and suggest they should carry on. In response to this, another participant suggested this could be addressed by having the robot contact the therapist directly to facilitate this exchange as necessary:

[P5]: "*you've also got to think about the questions that come along the route... the robot's just going to say 'stop you need to refer back to therapist', that sets up quite a sort of big message in the patients' brain of 'actually no I shouldn't be doing this because it hurts' and we might say 'oh its fine as long as its not making too much trouble for you, that's to be expected'..." [SL3]: "But then you could have the robot doing that, so it could say 'ok sort of keep going, keep going for the moment, maybe not quite so strongly, let's message through to the therapist now'...and if the therapist gets a text that says this is what's going on and she can give you a call... it could be worked in"*

Generally, participants who were less accepting or enthusiastic about the idea of using SARs were still able to identify ways in which they could be helpful; however these were more focused on fitting into current (conventional) therapy delivery. For example one participant expressed multiple doubts about robot technologies being suitable for occupational therapy (e.g. in their ability to 'read' the patient or be empathetic); however they were still able to identify how a robot might help to make better use of their own time with the patient:

[OT1]: "*I think you know the robot could help with something so if you say to the patient for next time prepare a list so that we can think about the groceries you need to do to be able to cook this meal or whatever it is I guess you know a robot or whatever could say have you prepared the list have you done this for tomorrow... so that sort of prompts that engagement into the task."*

In contrast, those that were more enthusiastic were able to come up with completely new applications and potential uses which the research team had not considered:

[SL1]: "*it's prompted me thinking of using some aspects of it for training other carers... a lot of the work I'm doing at the moment is teaching people in care homes working with people with dementia how to communicate with them... if you could show them how that works with a robot so 'this is what the robot is doing, isn't that more endearing when it does that, what would happen if you do that?'*

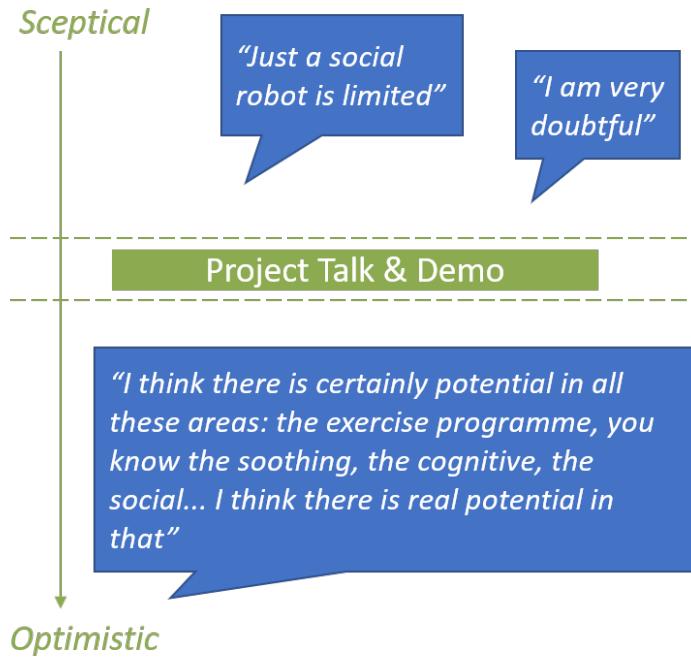


Figure 5.1: Example shift in comments made by one participant in the pre and post demonstration discussions.

5.2.3 Mutual Shaping: Impact on Participant Acceptance

As previously described, this focus group methodology is designed to support mutual shaping in three ways, as part of an overall mutual shaping approach to the design and deployment of robot systems. This includes mutual shaping that actually occurs *during* the focus groups, specifically the potential to have an impact on participants acceptance of robots. This was evidenced, both qualitatively and quantitatively, in the study with therapists.

Qualitatively this was evidenced by e.g. increasing positive comments and contribution to discussion. An example of this shift for one participant, who seemed to move from being very skeptical to more optimistic during the session, is given in Figure 5.1. In-group discussion analysis identified that a concern or issue raised by one participant was sometimes responded to or addressed by another without moderator intervention. This was observed at least once in each focus group.

[OT2]: “It doesn’t go upstairs does it... so you’d have to have another one upstairs” [OT3]: “Potentially but that’s something you can get over isn’t it” [OT2]: “But I mean it is a real issue then if someone’s downstairs during the day... then they’re upstairs for night time” [OT3]: “I’m sure there is an adaptation you can stick on the bottom of the robot that she could get up the stairs”

The difference in participant acceptance before and after the study is also evidenced by the results of a robot acceptance questionnaire (provided in Appendix A) administered pre and post-focus group. The questionnaire results were quantified (using reverse coding where applicable)

Statement	Result
Apprehensive	(Z = -2.818, p = 0.005)
Intimidating (to Me)	(Z = -2.309, p = 0.021)
Intimidating (to User)	(Z = 1.811, p = 0.70)*
Good Idea	(Z = -2.46, p = 0.014)
More Engaging	(Z = -3.116, p = 0.002)
Useful	(Z = -3.0, p = 0.003)
Improve Success	(Z = -2.48, p = 0.013)

Table 5.2: Wilcoxon Signed Rank Test results (N = 19) comparing pre and post-session acceptance questionnaire; all except * showing a significant increase in acceptance.

and a Wilcoxon Signed Rank Test was applied in order to measure the difference between the pre and post data sets for each acceptance statement. All statements except '*I think social robots might be somewhat intimidating to service users*' showed a significant ($p < 0.05$) increase in positive responses. The full results are listed in Table 5.2. The shift in acceptance is also shown visually for each statement in Figure 5.2, which shows the spread of individual participant responses, and Figure 5.3, which shows the shift in mean response across participants.

5.3 IML as Participatory Design for Mutual Shaping

5.3.1 How: IML versus Heuristic Implementation of an Autonomous SAR

The interactive, expert-in-the loop machine learning approach to automation was introduced and explained previously in Chapter 4. Here, it is demonstrated how that approach (i) fits in to a generalisable, end-to-end design, automation and deployment/evaluation process and (ii) compares to the use of expert designed heuristics as an alternative way to automate robots using expert input. Figure 5.4 presents both of these approaches in parallel, demonstrating the similarities and differences between them. This work compares both the practicalities of each approach as well as the autonomous systems resulting from them. For the heuristic implementation, that is design and testing of an expert generated rule based system, resulting in a fully autonomous system ready for deployment. For the IML implementation, that is design of the naive system and teaching interface, then training of that system until it is capable of running autonomously. Both processes were employed in parallel during the C25K gym study, to design two alternate versions of the C25K gym coach. Both versions of the robot were then tested with participants in order to allow for (i) comparison of their behaviour and (ii) within-subject participant/fitness instructor evaluation of the two systems.

Co-design of the Base System

Both methods begin with co-design of the 'base system'. This base system is the first version of the robot that will be deployed in the wild for user testing. The first step in the design process

5.3. IML AS PARTICIPATORY DESIGN FOR MUTUAL SHAPING

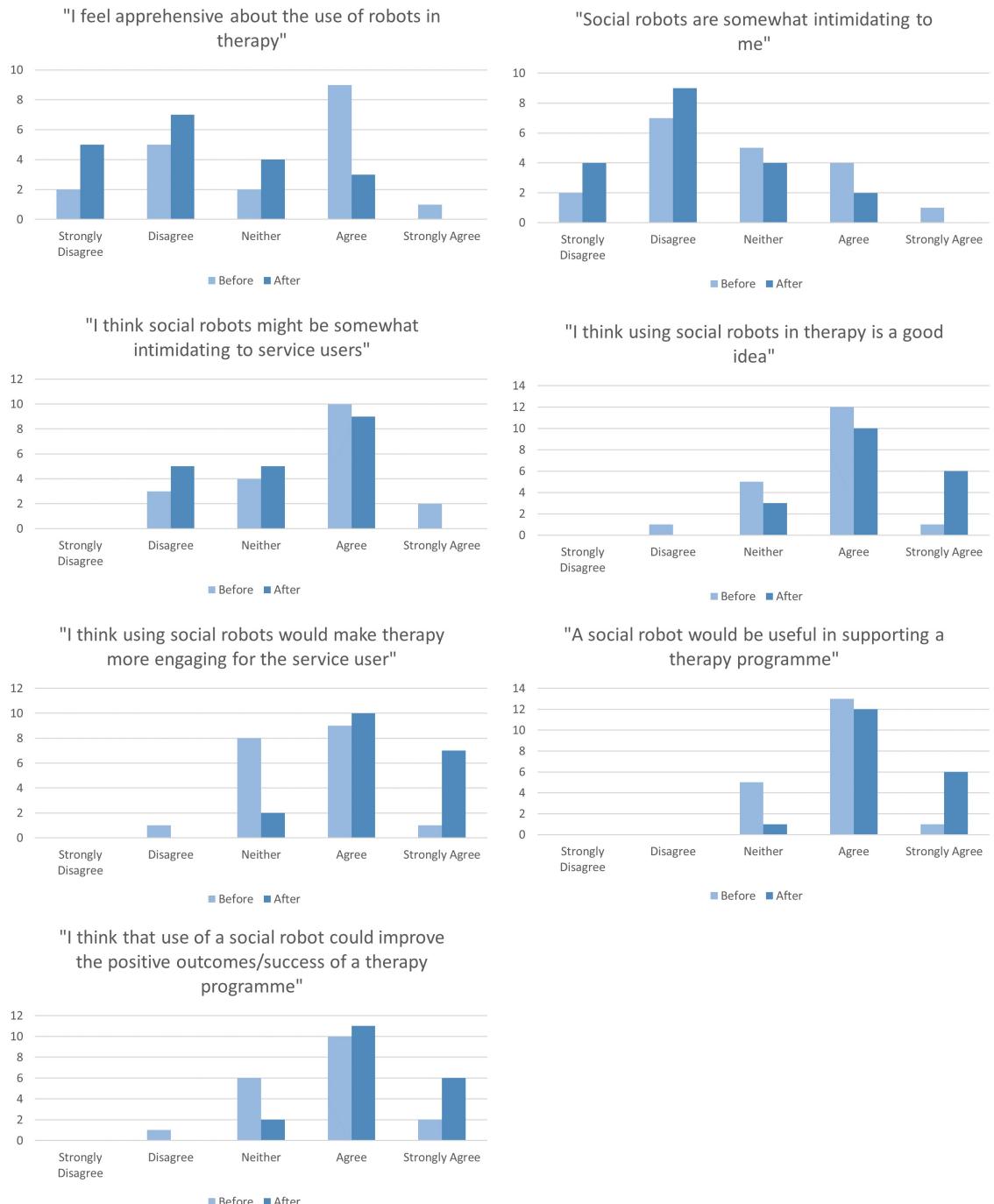


Figure 5.2: Distribution of participant responses ($N = 19$) to each question on acceptance of robots. For each question, the number of participants choosing each answer option is given for the pre and post-hoc questionnaire, administered before and after the focus group.

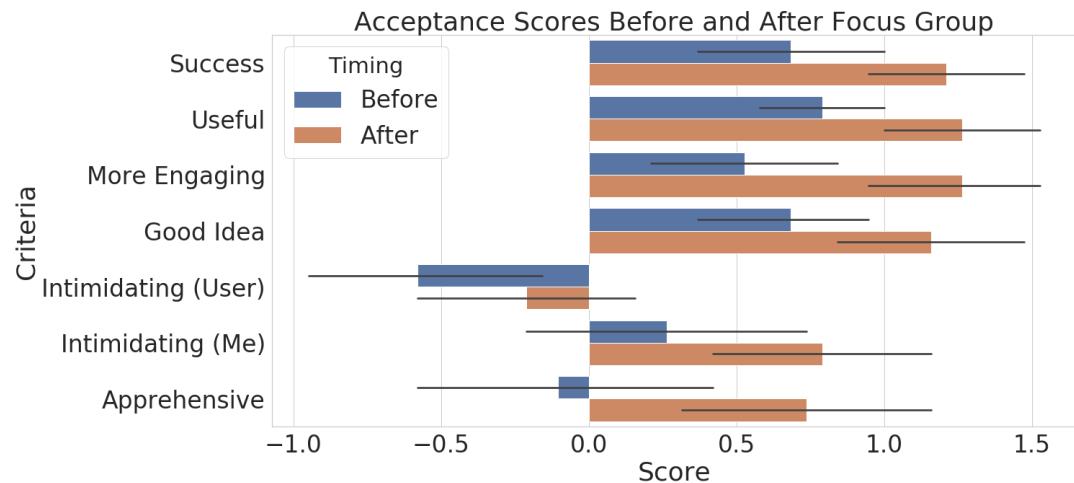


Figure 5.3: Mean participant acceptance scores with 95% confidence interval for the pre and post session questionnaires as another depiction of the shift between them ($N = 19$). Numerical scores represent coding of categorical answer choices shown in Figure 5.2, centered around 0 (neither agree nor disagree) and reverse coded where necessary to always reflect acceptance (i.e. positive score on ‘apprehensive’ = did not agree).

is introduction of the interaction scenario and proposed robot use case. This should be done at a relatively abstract level, such that participants i) can shape and refine the use case based on their expertise and experience and ii) can make recommendations/generate ideas unbiased by the researchers’ preconceptions. The research team can then present more robotics-specific related research, e.g. findings from previous work or related literature etc., before these ideas are revisited, as per the focus group methodology discussed under Section 5.2. At this point, the design team (researchers plus domain expert co-designer(s)) can start to draft the robot action and input spaces.

The action and input space should then be refined through an iterative process which includes physical prototyping (e.g. testing particular robot behaviours) and observations/testing in the context of use (i.e. running mock interactions with the expert in the the normal setting - see Figure 5.5). As discussed in Chapter 4, for the couch to 5km robot coach, this design work consisted of 6 co-design sessions conducted over a period of 5 weeks, representing a total of 12.5 hours direct co-design work. The heuristic based and naive learning systems were designed concurrently during these sessions. Resulting design of the heuristic system is presented in Section 5.3.1.1 and compared to that of the naive IML system in Section 5.3.2.

Throughout this process, the role of the domain expert is to help identify *what* actions the robot should be able to do (e.g. ‘ask participant how they are’) and *what* information might be relevant to informing action choice and content (e.g. running speed, heart rate). The role of the robotics researcher is then to consider more the *how* given e.g. the sensing and interaction capabilities of the proposed robot system. However these roles should not be fixed in an ‘expert

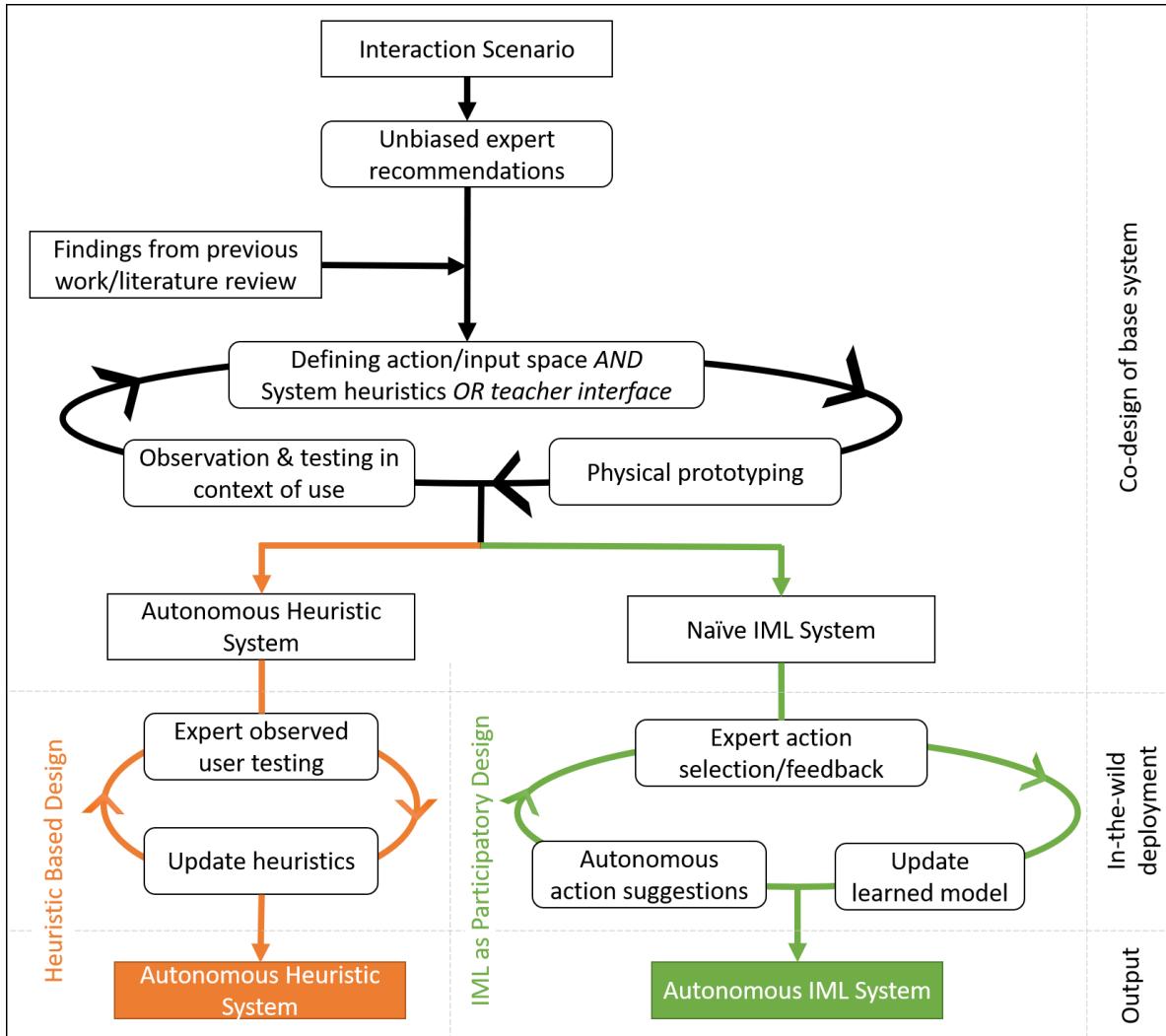


Figure 5.4: Alternative, generalisable, end-to-end participatory design and evaluation processes as employed for Heuristic and IML versions of the C25K robot coach developed during this work. Interactive machine learning as participatory design offers an alternative to using heuristics to encode domain expert knowledge in autonomous systems. In-the-wild deployment and iterative updating means that both processes can support mutual shaping to some extent, but the IML process allows the expert to tailor automated robot behaviour, in real-time, according to their tacit understanding of the situation.



Figure 5.5: Fitness instructor putting a colleague through a mock Couch to 5km style session for exploration and demonstration of instruction and encouragement actions.

wants x but roboticist can only provide y' dynamic. Rather, all participants should be actively working together to understand what's wanted and what's possible; reflecting the equal authority of all participants that is key to participatory design.

Regardless of the approach to autonomy, key outputs from this part of the process include:

- Well-defined interaction scenarios regarding the role of the robot (updated to reflect expert insight) and proposed 'in-the-wild' experimental design for IML automation and user testing
- Robot action space (base system outputs): a database of pre-programmed robot actions, designed with/signed off by the domain expert(s)
- Robot input space (base system inputs): identification/ implementation of required inputs, including any external sensors providing additional data input

For the heuristic based approach, rules for generating specific actions based on the available inputs should also be produced at this point. The result is a fully-autonomous system ready for deployment.

Design of the naive IML system requires identification of the action and input space only, nothing about the rules that link them. However, additional design activity is required to produce a teaching interface. It is through this interface that the expert will control and teach the system during the learning and evaluation process. As such, it is sensible to include design of a teaching interface in the participatory design process to ensure that the interface:

- provides the expressivity required for the expert to specify the social behaviours (action type, content, timing etc.) to be taught.
- is intuitive when used in real-time within the context of use.



Figure 5.6: Fitness instructor working on physical prototyping of the teaching interface.

For the C25K robot coach, this was achieved through paper prototyping (see Figure 5.6) and testing of a draft implementation in a mock session, concurrently with action design, early in the process.

For the IML approach, the result of this initial co-design process is the naive robot and teaching system, made up of the following key elements:

- Robot with output actions and data input as per the base system (although likely to have increased action/input space compared to heuristic system as per Tables 5.3 and 5.4)
- Naive machine learning agent ready to accept training examples
- Teaching interface through which the expert can provide teaching examples/feedback to the learning system

As shown in Figure 5.4, both approaches can be implemented through an iterative process which includes prototyping and testing in the context of use. Testing in the context of use is particularly important because it can be quite difficult for professionals from human-centered domains (e.g. health, education) to explicitly identify their reasoning in taking particular actions.

In-the-Wild Testing and Evaluation

In-the-wild user testing and evaluation is a key part of the proposed design process, regardless of the approach to autonomy. However, as shown in Figure 5.4, its role in the overall design process also represents where the two approaches really diverge. Details on each are given below.

In both cases however key requirements for this testing, in order to support a mutual shaping approach to robot design, are as follows:

1. Should be carried out in contextually appropriate environment (e.g. an actual gym for the C25K robot)
to facilitate accurate rendition of natural and appropriate behaviours from the expert / accurately assess performance of the system
2. Should be longitudinal, i.e. involves multiple participants for a significant period of time
to overcome inter-participant variability and novelty effect, and to allow for sufficient generation of training data / iterating of heuristics
3. The domain expert should be present at all sessions
expert maintains active role in teaching and / or iteratively updating the system; also provides a constant source of expert system evaluation
4. Recruited participants should represent realistic end-users of the system (not a convenience population)
again to facilitate rendition of appropriate behaviours as well as minimise bias and maximise potential for generalisation of the resultant system

For the IML approach, how in-the-wild experimental time is ‘split’ between training and testing/evaluation, as well as the choice of evaluation measures, will depend on the use case and research questions of interest. However, the continual presence of the domain expert allows for continued participation in evaluation of the system; so they should document observational notes throughout. The result is an autonomous system, trained by an expert (and with the potential for continual expert improvement) as per the C25K robot in Chapter 4.

For the heuristic approach, observations made by the expert co-designer can be used to update the heuristic control logic. Ideally, user feedback should also be considered at this stage. Depending on the application and stage in development, this could be done via the expert co-designer (i.e. their observations including user feedback) or through inviting participants to become co-designers of a next iteration of the system. This should be repeated iteratively until all stakeholders are satisfied with the resultant system. The following subsection demonstrates the resultant output of this process when applied to the C25K scenario, as a control/comparison for the IML system presented in Chapter 4.

5.3.1.1 A Heuristic Based Couch to 5km Robot Coach

Design of the heuristic system was undertaken directly in parallel to design of the IML system, as per the methodology presented in Figure 5.4 and the design activities highlighted in Table 4.1 of Chapter 4.

The overall motivation behind the design of the heuristic system was that it should:

- (i) challenge them if their speed was at 50% or below of their average walking/running speed throughout the programme so far
- (ii) reward them if they reached a new personal best speed
- (iii) sympathise if their speed was in-between those two speeds described above and their heart rate was at or above 80% their maximum recommended exercising heart rate
- (iv) otherwise, if participants' weren't in one of the above categories, tell them to maintain their current performance

Iterative updating of the system between Phases 1 and 3 of the study resulted in addition of the following:

- (v) ask participants how they are feeling half way through the run
- (vi) tell participants to slow down if their heart rate exceeded their maximum exercising recommended heart rate

Algorithm 4 shows the resultant algorithm for producing the H robot's autonomous behaviour. Notably, this resulted in a reduced action and input space compared to the IML system, demonstrated by Tables 5.3 and 5.4 respectively. Even after iterative updating of the heuristics between Phase 1 and Phase 3, the final heuristic control algorithm only utilises 6 actions, compared to the 12 actions available to the IML system. The fitness instructor was simply unable to extend the heuristics further, to include actions such as *humour*, *get closer* or *speed up* for example because:

- no obvious, universal conditions for using such actions could be identified
- such actions were considered 'risky' by the fitness instructor; i.e. telling a joke at the wrong time could severely negatively impact on credibility of the robot and/or user experience
- in the case of speed up, incorrect usage could potentially be unsafe

A similar case can be made for the comparatively reduced input space shown in Table 5.4. For example, the fitness instructor identified that participant personality would be an important factor in deciding how best a human and/or robot fitness instructor should try to motivate them. However, he was not able to identify any way of explicitly linking e.g. personality type or personality scores to specific use of the robot's action space. This exactly echos results concerning personalisation of therapist behaviour in Chapter 2. Therapists were similarly unable to identify personalisation heuristics, even when working with an NHS-designed categorisation tool for describing people's attitude to healthy living and informing approaches to motivating/persuading them. A key benefit of the IML approach is that it addresses exactly this - the expert can identify such inputs *without* needing to also identify *how* they should inform system behaviour.

Action	IML-S	IML-A	Heuristic
Challenge	✓	✓	✓
Praise	✓	✓	✓
Sympathise	✓	✓	✓
Maintain	✓	✓	✓
Speed Down	✓	✓	✓
Check User Exertion	✓	✓	✓
Time	✓	✓	x
Humour	✓	✓	x
Animation	✓	✓	x
Get Closer	✓	✓	x
Speed Up	✓	x	x
Back-Off	✓	x	x

Table 5.3: Comparison of the action space available to the IML and Heuristic systems, the IML-S column shows that all actions available to the IML system were executed and utilised by the fitness instructor (in IML-S sessions). The IML-A column shows which actions were executed by the IML system when running autonomously (in IML-A sessions).

Input	ML	Heuristic
Time since last action	✓	✓
Speed	✓	✓
Heart Rate	✓	✓
Session progress	✓	✓
Facial expression	✓	x
Time	✓	x
Programme progress	✓	x
Personality traits Gosling et al. (2003)	✓	x
Activity level	✓	x
Attitude to exercise	✓	x

Table 5.4: Comparison of input spaces for the IML and Heuristic systems.

Algorithm 4 Final heuristic based control algorithm for Heuristic robot system, developed initially during pre-study co-design and then updated ahead of Phase 3 testing based on the fitness instructor's observations during Phases 1 and 2. The lines marked * were additions made during this update.

Input: State x

```
if x.session_progress == 0.5* then:  
    next_action = [SYMPATHETIC, CHECKPRE]  
else if x.heart_rate >= 220-age* then:  
    next_action = [SYMPATHETIC, SLOWDOWN]  
else if x.speed <  $\frac{\text{average\_speed}}{2}$  then:  
    style = random_choice[SYMPATHETIC, CHALLENGING]  
    next_action = [style, CHALLENGE]  
else if x.speed >= personal_best then:  
    style = random_choice[SYMPATHETIC, POSITIVE]  
    next_action = [style, praise]  
else if  $\frac{\text{average\_speed}}{2} < x.\text{speed} \leq \text{personal\_best}$  and x.heart_rate >=  $0.8 \times (220 - age)$  then:  
    next_action = [SYMPATHETIC, SYMPATHISE]  
else  
    next_action = [POSITIVE, MAINTAIN]
```

5.3.2 Why: Benefits of Process and Resulting Autonomous System

The results in Chapter 4 demonstrated the successful use of IML to create an effective SAR that produced appropriate autonomous behaviour. However, to further demonstrate the benefits of this approach specifically, the resulting system was also compared against the heuristic based autonomous system described above. Within-subject comparison of the two systems was therefore incorporated into the overall experimental schedule, specifically in Phases 1 and 3 of the study as described in Chapter 4. As a reminder:

- In Phase 1 [8 sessions per participant]: Participants alternated between (a) the IML system as supervised/ultimately controlled by the instructor (IML-S) and (b) the first iteration of the H robot (H1) each session, for a total of 4 sessions with each. It was made explicitly clear that the two robots were programmed differently (each robot was colour coded either orange or purple) but not it was not explained how they were different.
- In Phase 3 [2+ sessions per participant]: Participants unknowingly worked out with IML-A robot (i.e. they were not briefed about the change in robot control in any way) for two sessions before the updated H robot (H2) was explicitly re-introduced for a single session. The difference between the systems was hidden as previously.

As such, all participants worked out with the H robot for a total of 5 sessions. Specifically concerning comparisons between the IML and H robots, Phase 1 was designed to allow within-subject testing of the IML-S versus H robot, and Phase 3 was designed to allow within-subject

testing of the IML-A versus (updated) H robot. Detail on the experimental measures designed to capture participant experience of the programme and various robot conditions are described in Chapter 4, and given in full in Appendix C.

5.3.2.1 Hypotheses

H1A The H system will produce behaviour very different to that of the IML-S/A systems (even though it was designed by the same fitness instructor).

H1B The H system will not demonstrate personalised behaviour across participants.

H2 Participants will evaluate the the IML-S robot more positively than the H robot.

H3 Participants will evaluate the IML-A robot more positively than the H robot.

H4 The fitness instructor will evaluate the IML-A robot more positively than the H robot.

Note there is a subtle link between H2 and H3 regarding IML as a process in general versus the work undertaken in the C25K study as specific implementation of that process. If demonstrated to be true, H2 provides the motivation for attempting this IML approach when considering SAR automation. H3 is then more a statement regarding how well the specific implementation of this process, undertaken for this work (as presented in Chapter 4), was able to deliver on that motivating factor.

5.3.2.2 Results

For these results, qualitative data regarding participant evaluation of each robot are taken from questionnaires and post-session/weekly feedback collected throughout the study. Quantitative data regarding the actions executed by each version of the robot (to compare for similarity/differences in robot behaviour) are taken from a subset of the most comparable sessions. Table 5.5 identifies what session data were used to compare behaviour of the heuristic, supervised and autonomous robot behaviour. The performance of the heuristic robot is compared to that of fitness instructor supervised behaviour at two stages of the experimental study: Phase 1 (IML-S1 and H1) and the end of Phase 2 (IML-S2 and H2). The former analysis is based on 36 sessions: 2 heuristic and 2 supervised sessions per participant whereas the latter is based on only a single session per condition per participant. Comparing these differences in Phase 1 and Phase 2 separately allows for comparisons between (i) behaviour of the heuristic system before and after it was updated and (ii) behaviour of the supervised system fairly early on in the programme versus later, when the fitness instructor may have been expected to understand participants' differing needs better as he observed them training. The performance of the heuristic system is also compared to that of autonomous behaviour demonstrated in Phase 3 of the experiment (H2 and IML-A) based on comparisons across a single session per condition per participant. The

Participant	Phase 1 Comparison		Phase 2/3 Comparison		
	IML-S 1	H1	IML-S 2	H2	IML-A
LB	2,4	3,5	22	25	24
FB	2,4	1,3	21	24	23
DB	1,3	2,4	17	21	20
JF	1,3	2,4	22	25	24
MR	2,4	1,3	18	21	20
DP	1,3	2,4	24	27	26
JW	5,7	2,4	22	25	24
GB	2,4	1,3	23	26	25
PT	2,4	1,3	18	21	20

Table 5.5: Subset of sessions used to compare H, IML-S and IML-A robot behaviour.

compared sessions have been selected to be as similar as possible with regards to e.g. session length and difficulty after accounting for any technical difficulties that prevent particular session data being used.

Resultant Robot Behaviour

The distribution of (*action-type* and *action style*) of actions executed by each version of the system are shown in Figures 5.7 and 5.8. Figure 5.7 also demonstrates the impact of the reduced action space available to the heuristic system, as presented in Table 5.3 and discussed in Section 5.3.1.1. Fisher's exact tests were applied to these distributions within-participant, to test whether each version of the robot produced significantly different behaviour for the same participant. For all 9 participants, the heuristic system produced behaviour significantly different to that of the IML system, both as supervised in Phases 1 and 2, and when running autonomously in Phase 3. In addition, they demonstrated that the updated Heuristic system (H2) behaved significantly different than the first iteration (H1) for all participants. The full results table is presented in Appendix D.

As discussed in Chapter 4, it should be noted that comparing overall action distribution is only a *partial* measure of how similar the sessions are (as it doesn't account for e.g. timing of actions within the session). Further such a comparison *does not* give any indication of how *appropriate* those (dis)similar distributions might be, given that two 'good' sessions, where the robot acts appropriately, may have very different action distributions. The results demonstrate:

- (i) The heuristic robot produced significantly different behaviour to that of the IML system (both when supervised and when autonomous) across all participants.
- (ii) The updated heuristic robot (H2) produced significantly different behaviour to that of the initial heuristic robot (H1) across all participants. This demonstrates that the fitness

instructor-led update to the rule based driving the Heuristic system, implemented between Phase 1 and Phase 3, had a significant impact on the system's resultant behaviour.

These results provide strong support for hypothesis H1A: *the H system will produce behaviour very different to that of the IML-S/A systems (even though it was designed by the same fitness instructor)*. This is likely due, most of all, to the difference in action spaces available to the robot. As discussed previously, the Heuristic system had a smaller action space than the IML system due to the fitness instructor being unable to generate heuristics for utilising some of the more complicated social actions.

The distribution of (*action-type* and *action style*) of actions executed by each version of the system are shown *between participants* is shown in Figure 5.9. Fisher's exact tests were applied to these distributions between-participants, to test whether either iteration of the Heuristic system produced significantly different behaviour across participants. For the first iteration H1, there were only 5/36 pairwise participant comparisons that were **not** significantly different. For the second iteration H2, this increased to being 11/36 pairwise participant comparisons that were **not** significantly different. The full results table is presented in Appendix D.

These results suggest that actually the heuristic based system did result, on the whole, in very different behaviour across participants. As such, there is strong evidence **against** hypothesis H1B: *the H system will not demonstrate personalised behaviour across participants*. The input space for the Heuristic system means these differences can only be driven by dynamic performance of the individual (with respect to their previous performance) and so therefore presumably simply reflect variation in participants' efforts and relative performance in those sessions selected for comparison.

Participant Evaluation

Figure 5.10 shows participant responses to questions on which robot (if any) (i) they performed best with, (ii) they preferred and (iii) they would prefer to work with in future, collected at the end of Phase 1 testing and again during Phase 3 testing. At both of these timing points, the IML robot was the most commonly given answer across all three measures. However, the results were not unanimous, and the distribution of answers did change between the Phase 1 and Phase 3 testing. In the Phase 3 questionnaire, participants were also asked to describe both the IML and H robots as fitness instructors, and any perceived differences between them. Their answers are provided in full in Appendix D.

At both times of testing, 3/9 participants felt they performed best with the heuristic system, although only 1/3 of those participants (LB) gave this answer at both Phase 1 and Phase 3. LB did not provide any reasoning for this choice at either stage of testing. The other two participants who felt they performed better with the heuristic robot in Phase 1 expressed an overall preference for the heuristic robot across all of the measures. For DB this appeared to be a clear preference:

5.3. IML AS PARTICIPATORY DESIGN FOR MUTUAL SHAPING

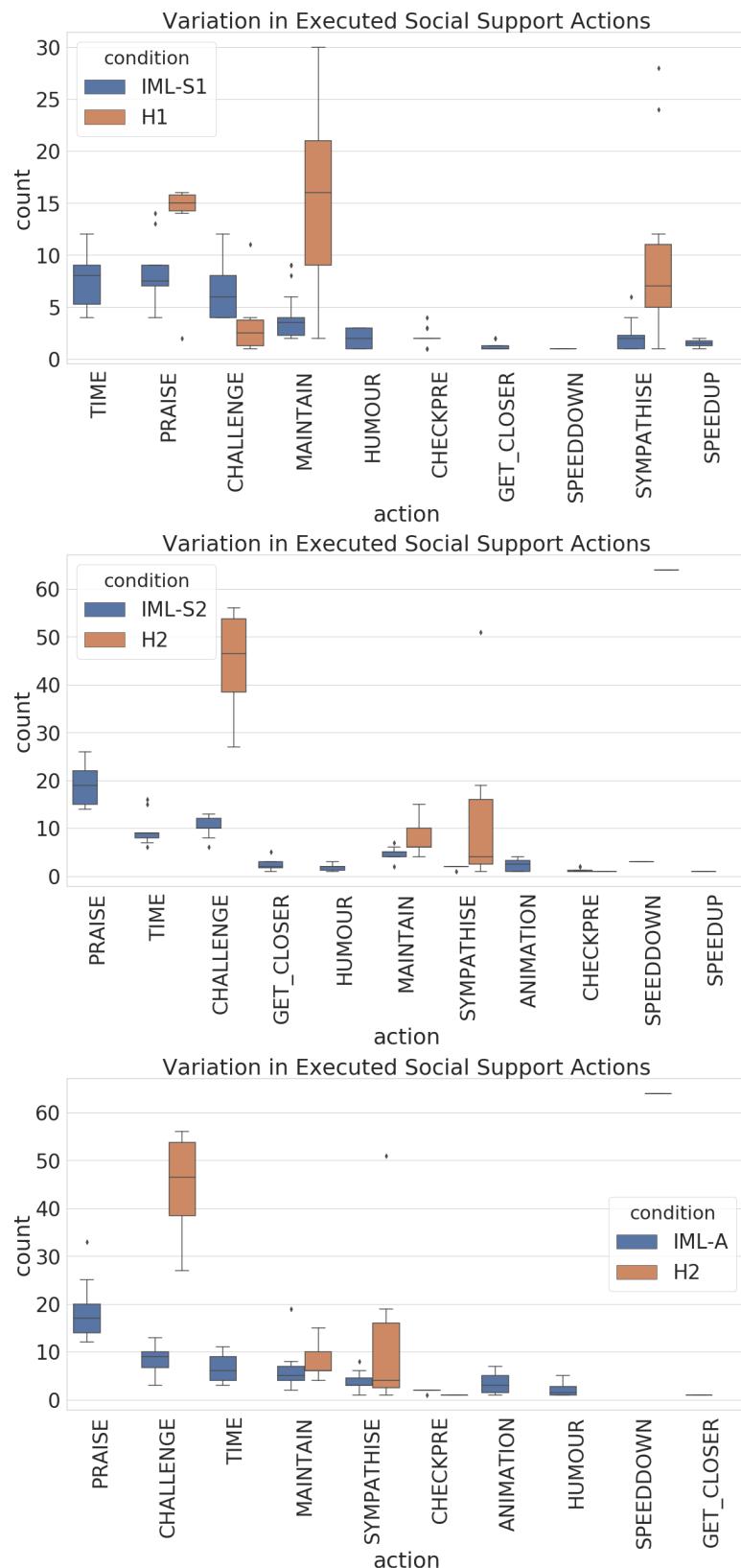


Figure 5.7: Comparison of action-types produced by supervised, heuristic and autonomous robots.

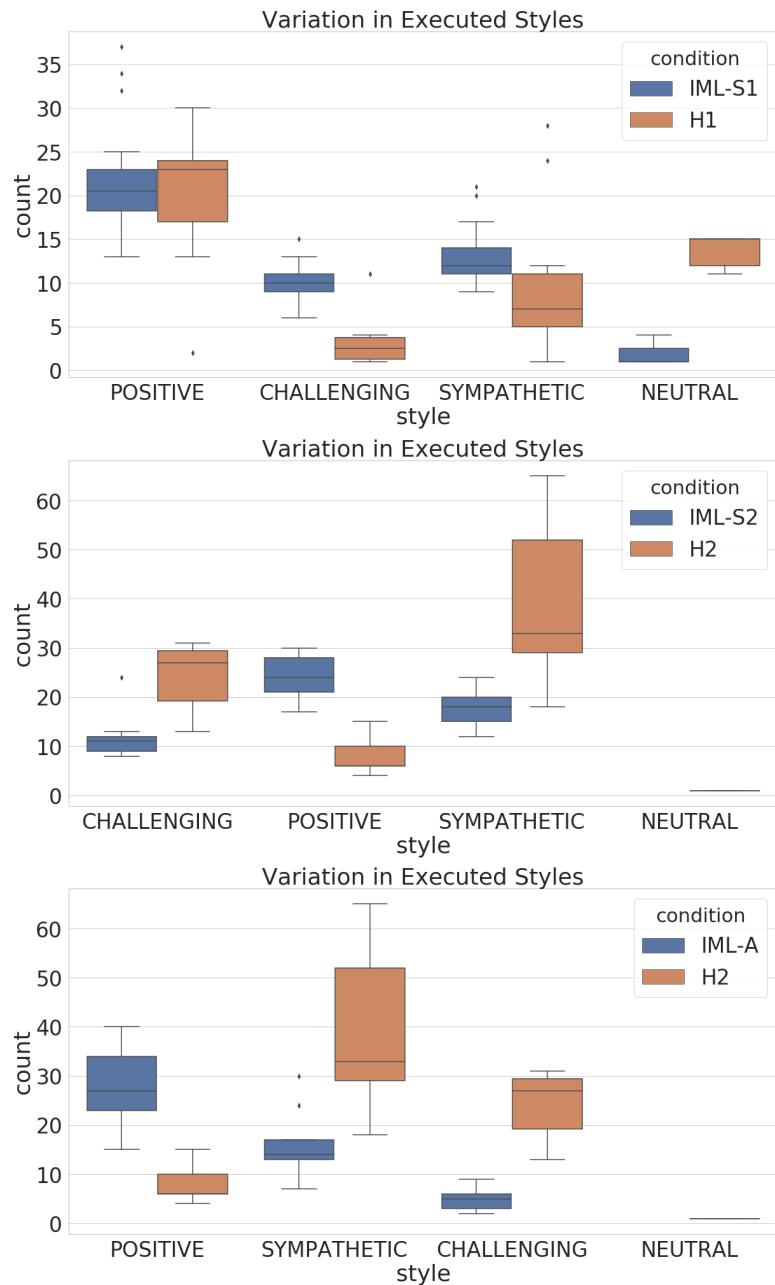


Figure 5.8: Comparison of action styles produced by supervised, heuristic and autonomous robots.

5.3. IML AS PARTICIPATORY DESIGN FOR MUTUAL SHAPING

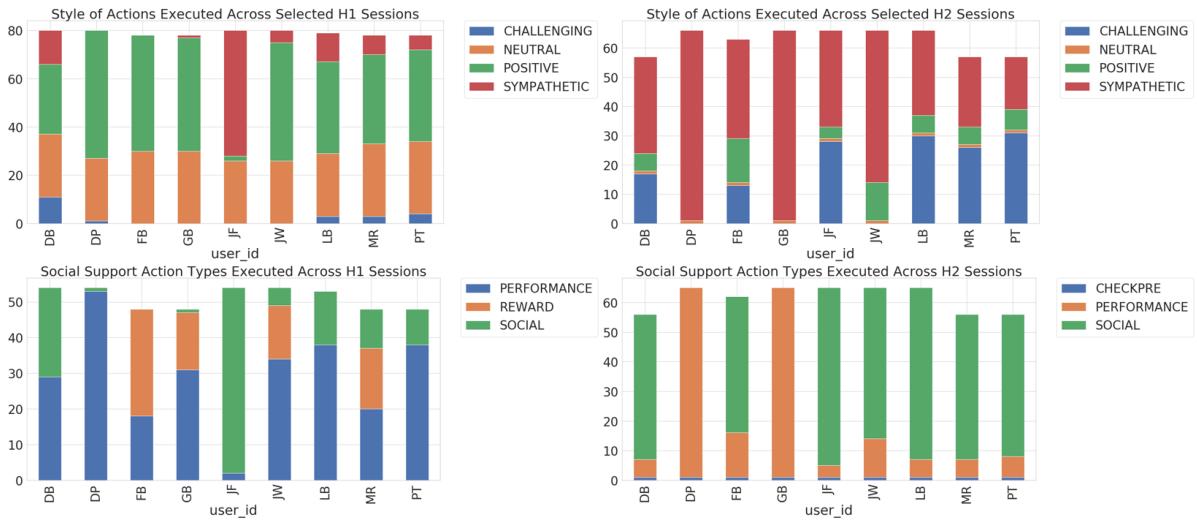


Figure 5.9: Style and action-type of actions executed by the heuristic system across participants.

[DB]: '*[H] was nice and encouraging and got the balance right; [IML] is annoying...too much info'*

whereas DP suggested he saw little difference between them:

[DP]: '*I do not really notice the differences*'.

At Phase 3 testing, the other two participants who felt they performed better with the heuristic robot MR commented this was '*only a little*', and responded with '*no difference/preference*' to the other two questions, whereas for JF, this seemed to be linked with her perception of the differences between the two systems:

[JF]: '*I think the [H] robot focused more on trying to get people to put in the most or more effort whereas the [IML] robot was more gentle and wanted people to just try the best they could*'.

This was also reflected in their answers to the questions on which robot they preferred (no preference):

[JF]: '*I think it would depend on where I was in the training programme and what type of day I was having*' and would rather work with in future (H): '*I'd want to push myself now and this one would be better though [IML] better on days feeling a bit weaker*'.

Comparing individual participants' questionnaire responses at Phase 1 and Phase 3: 5/9 participants gave different answers to all three questions, 2/9 participants gave different answers to two questions, 1/9 participants gave a different answer to a single question only and 1/9 participants did not change their answers at all. Figure 5.10 shows this manifested as more participants having less clear preferences (i.e. selecting 'no difference') in the Phase 3 questionnaire. Potential reasons for this shift could be:

1. Autonomous behaviour of the IML robot had a negative impact on participants' preference for it/perception of it, thus reducing preference compared to Phase 1 testing.

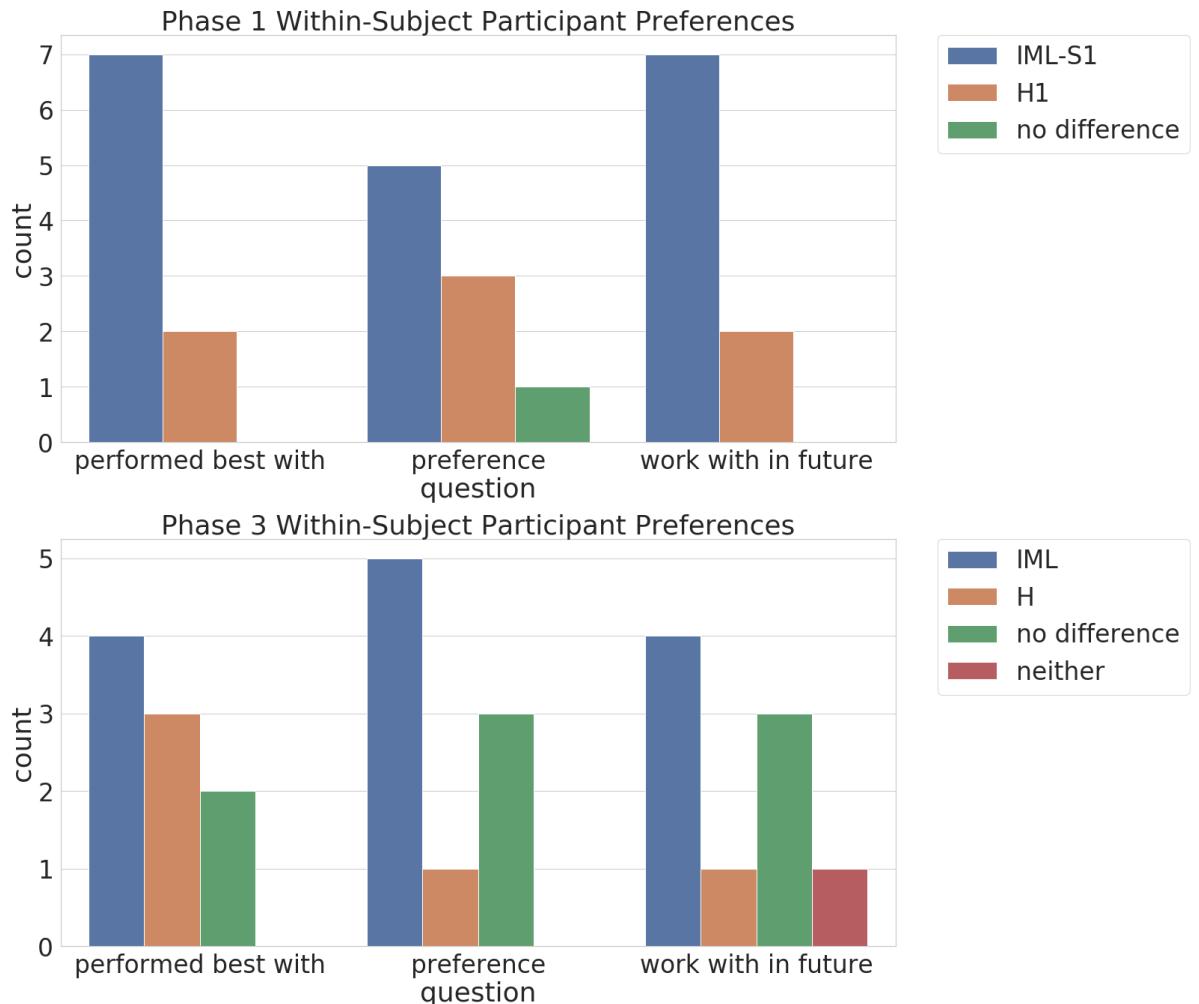


Figure 5.10: Participant preferences regarding a) the Heuristic and IML-S robots at the end of Phase 1 and b) the Heuristic and IML(-S/A) robots after at least 2 sessions with the IML-A robot at the end of Phase 3. Whilst the aim of the secondary round of questions was ideally to have participants compare the Heuristic and IML-A systems, it cannot be guaranteed that participants were not drawing on their *overall* experience with the IML system (i.e. including those sessions where the robot's behaviour was ultimately controlled by the fitness instructor). Figure shows responses to questions on which (if either) of the two robots i) encouraged them to perform better, ii) they preferred and iii) they would prefer to work with in future, collected at the end of Phase 1 (after alternate testing of the H and IML-S robots) and during Phase 3 (after at least two sessions working the IML-A and one session with the updated Heuristic robot).

As discussed in Chapter 4, on making the switch from supervised to autonomous operation of the IML robot, only 2/9 participants directly noted a negative change in their post-session feedback. 4/9 participants gave no indication that they noticed any change at all. Similarly, other than DB who expressed an active dislike of both robots/preference not to work with any robot again in future, no participants left any negative feedback concerning the IML robot during the end of experiment measures. The same cannot be said for the Heuristic system, for which e.g. GB and DP both left negative feedback after Phase 3 testing - discussed further under point 2.

2. There was less perceived difference between/impact of IML-A robot versus Heuristic robot behaviour (following the switch to autonomy and iterative updating of heuristics between Phases 1 and 3).

Participant feedback suggests 8/9 participants perceived some identifiable difference between the heuristic and IML robots during Phase 3 testing, suggesting they were distinguishable. However, the nature of these comparisons varied across participants based on the specific behaviour they had seen, in some cases (as previously discussed for JF) these differences also represented participant reasoning for having no preference between the two systems. 3/9 participants documented strong negative reactions to the Heuristic robot specifically:

[GB]: *'I'm not a big fan of the [H] robot...it was quite annoying and difficult to run with. I didn't like the fact it was telling me to slow down from a pace that I was comfortable with, its repetitiveness and negativity was doing my nut in.'*

[DP]: *'[H] pepper was very different...I did not find it helpful'*

[DB]: *'I hate [the H] robot she is incredibly annoying'*

In addition, 4/8 participants who then proceeded to work out with the IML-A robot again, after working with the Heuristic robot, specifically praised it/identified as being happy to swap back, including MR who had previously explicitly suggested they had no preference and FB who didn't describe any significant difference when comparing the two robots:

[FB]: *'glad to have [IML] robot back today! loving the little dance moves at the end'*

[MR]: *'I hadn't realised how much the [IML] version of pepper actually helps me to keep a positive mindset'*

3. Effects stemming from the long-term nature of the study and/or complex testing schedule/experimental design.

The long-term nature of the study and the repetitiveness of interactions inherent to the couch to 5km programme make it likely that participants were no longer experiencing much, if any, novelty effect on working with the robot towards the end of the programme. Further, given that the robot was programmed with a limited set of speech utterances, participants became familiar with these specific phrases and were increasingly aware

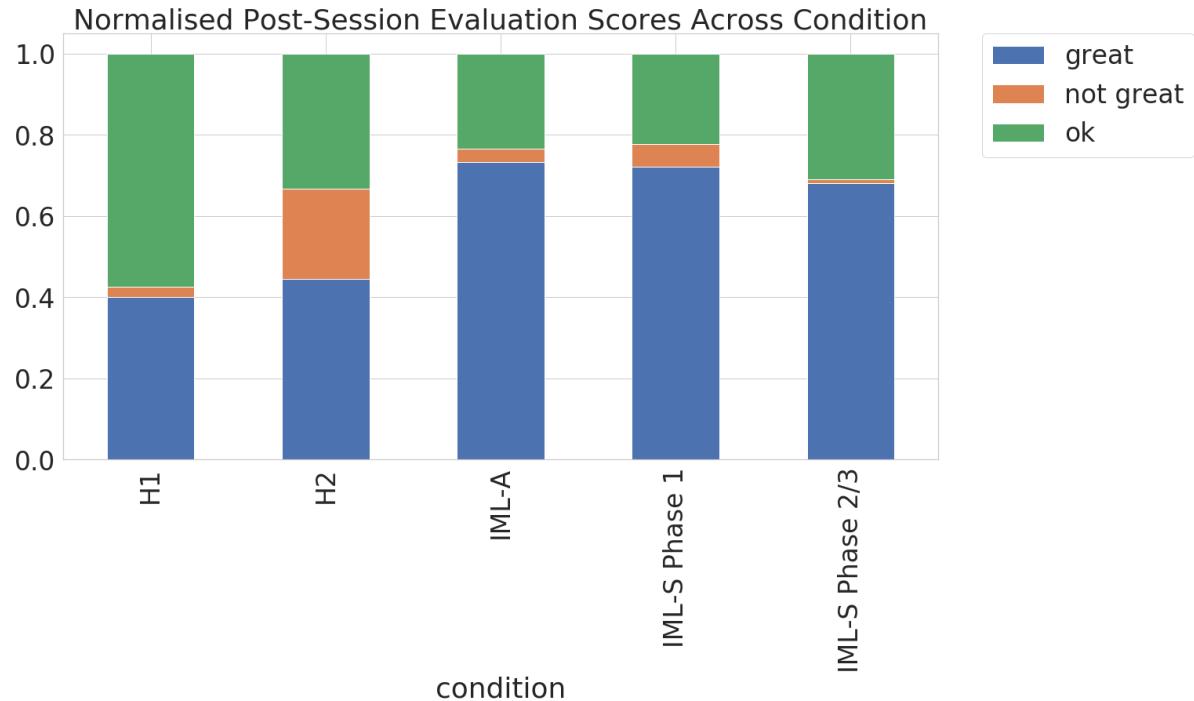


Figure 5.11: Participant post-session evaluation of the robot across conditions, normalised for the number of each session type.

of/somewhat fatigued by repetition in the robot's speech as the study progressed (evidenced by qualitative feedback in the post-session measures). This repetitiveness was common across all robot conditions, potentially therefore reducing the perceived difference between them:

[DB]: '*I don't tell the robots apart [H], [IML] - makes no difference, as the text is repetitive'* Compounding this, the Phase 3 experimental schedule was designed such that participants only saw the updated heuristic robot for a single session, meaning that participants may have been wary about dismissing it/did not interact with it long enough to have a clear preference (as they did in Phase 1):

[JW]: '*I didn't really have time to get to know [H] Pepper so I don't know which I'd prefer.'*

Figure 5.11 shows participant normalised evaluation ratings of each robot condition, as collected after every session. It can be seen that the H1 robot received a lot more 'ok' ratings than the equivalent IML-S at Phase 1 robot, and similarly the H2 robot received a lot more 'not great' ratings than the IML-A robot. Only the IML-S and H1 evaluation scores were significantly different however, as demonstrated by a Fisher's exact test returning $p < 0.01$.

Overall, these results provide moderate-to-strong support for H2 and H3: *participants will evaluate the the IML-S robot more positively than the H robot and participants will evaluate the IML-A robot more positively than the H robot as most participants did evaluate the IML system*

more positively than the H robot overall. However, they also demonstrate evidence for the idea that preferences for robot *style* or behaviour vary both between and within-participant, and that for one participant in particular there was simply no desire to really work with the robot at all.

This in itself is an unsurprising result, and completely consistent with previous discussion on (i) the need for personalisation and (ii) the recognition that there are some people for whom a robot would never be appropriate nor accepted. Further, these results reflect what might be expected for *any* fitness instructor, whether they be robot or human, in that different clients will prefer different instructors, potentially even on different days, based on how they are feeling etc.

Fitness Instructor Evaluation

Fitness instructor notes from Heuristic robot sessions in Phase 1 and Phase 3 testing (provided in Appendix D) were checked for references to the robots performance. The Phase 1 feedback is relatively mixed, with evidence of sessions made up of good/appropriate action choices but also where this was lacking. A common theme also appears to be the potential for repetition and lack of variety in the robot's actions in some sessions. The Phase 3 feedback is similarly mixed but arguably showcases much stronger criticism of the Heuristic system in particular sessions, specifically for participants DB, DP and GB. A common theme here appears to be the lack of evolution of behaviour/speech throughout the session, such that e.g. the robot does not offer more challenging actions near the end of the run as the instructor may do.

In a post-study interview, the fitness instructor was asked specifically to reflect on performance of the Heuristic system, as per the following extract:

[KW]: Final comments on the heuristic robot versus the learning one - how did you feel in the end? Looking back, you obviously were heavily involved in both of them. Any strong thoughts or feelings when you think back about how they both performed?

[Instructor]: *So obviously the heuristic was the best sort of hard-coded idea we could produce and it was, it really was. It was intelligently designed with good sort of specifications about the situation and the physical performance of what's going on, and it did do alright. Especially in those early runs when they were really short, and so like the frequency of the speech wasn't too bad. It was good. It's just when the runs started to get longer and the clients you know, were a few weeks into the programs so the clients already started to know Pepper's speech. That's when it started to just sort of fall short of expectations and just no longer became that sort of, that successful control and it started to show there was more sort of that robotic, repetitive. It's not really taking into consideration the full aspect of what's going on, the situation, whereas the with the learned A.I., it made some intelligent decisions. It like, oh, sometimes it was slow, but it made those decisions. And there are some some cases where I thought, you know, the client was really pushing themselves, they're finding it tough. And the robot asked how they were, which was the perfect time to answer that intelligent question. And it responded to the client's response and in the appropriate manner. It was really good. And I couldn't see the control making that same, like it couldn't even make that behaviour. So, um, yeah, I really saw a difference. And I could tell that that level of teaching or*

reinforcement really, yeah really did stick, really paid off.

Overall, these results provide strong support for hypothesis H4: *the fitness instructor will evaluate the IML-A robot more positively than the H robot.*

Other Benefits of the IML Process vs Heuristic Design

In addition to the increased action/input space, another key benefit inherent to the IML process but much less feasible for heuristic design, is the potential to reflect dynamic changes in desired behaviour (demonstrated in Chapter 4). This allows for changes in the instructor's use of the IML system based on e.g. increased understanding of the participant and/or participants needs changing as the programme difficulty increased. In the Phase 3 questionnaire, participants were asked whether they perceived a difference in either of the robot's behaviour over the course of the study. 5/9 perceived a change in the IML robot's behaviour. None of the associated comments were negative, nor mentioned behaviour changes that might be associated with switching from supervised to autonomous (e.g. increased repetition), but rather cited the an improvement in the robot's 'understanding' of them/ability to motivate them. In the case of FB, this was also their explicit reasoning for preferring the IML robot over the H robot:

[FB]: *'I didn't prefer the [IML] robot in the beginning because I thought it was a bit pushy. After the initial sessions, I felt it included more encouraging/supportive phrases and had a good balance with the phrases designed to push you. They were also well placed within the runs (often towards the end where you might need more motivation).'*

[PT]: *'I think now [the IML robot] is more informed and gives more accurate feedback (i.e. to slow down when I am really tired)'*

[JF]: *'[the IML robot] seemed like it got more helpful in getting me to find a flow'*

[GB]: *'It's hard to say whether the [IML robot] actually developed or whether my familiarity with it just improved, but it seemed like towards the end its commenting was at the right frequency, whereas earlier on in the program I thought it was commenting too much or sometimes too little. I also don't know if the types of comments it chose changed, but I certainly found what it said to me more useful at the end than it did at the beginning.'*

Such comments are further evidence that the fitness instructor did increasingly tailor the robot's behaviour based on his understanding of what worked well for these participants as he got to know them. This learning is reflected in notes he took for himself through the programme, e.g. *push her/keep her interested, challenge further, starts off way too fast* etc. and the personalisation of IML robot behaviour discussed in Chapter 4. Such personalised and/or programme specific changes are very difficult to implement into population-wide heuristics. Essentially, the IML process offers improved flexibility/ease of reacting to change when compared to heuristic design. This flexibility is specifically what gives the IML approach increased potential to support a mutual shaping approach to robot automation. Specifically, it allows for expert shaping of autonomous robot behaviour to reflect any desirable changes resulting from the manifestation of mutual shaping effects that only become apparent on robot deployment. GB's comment specifically also

identifies the potential for another type of mutual shaping effect that affects quality of interaction and hence effectiveness of the system; that of user familiarity with/adaptation to its use.

Finally, it seems likely that the active, real-time involvement of a domain expert required for this IML automation process would have a positive impact on (i) increasing acceptability of and (ii) reducing ethical risk associated with the resultant system. Results in Chapter 4 demonstrated that the instructor's presence throughout the study had a positive impact on participants' confidence in working with the system, and so it seems likely that making participants fully aware of his role in training the system would have a positive impact on acceptance of resultant autonomous behaviour. On ethical risk, the involvement of a domain expert who can tailor the robot's behaviour, in real-time, to a particular user's needs, and/or intervene to prevent particular behaviours being executed, offers both increased safety but also reduces the risk of dehumanisation associated with human-robot interaction (particularly where robots are guiding or instructing user behaviour) as described in BS 8611 BSI (2016).

5.3.3 Mutual Shaping: Interactions Between the IML System, Fitness Instructor and Participants

The real world deployment of the IML robot necessary for its automation also allowed for the observation of mutual shaping effects regarding its use. One particularly clear example of such, which emerged towards the end of the study, was the fitness instructor's use of the robot in the case of overlap between participants. As the study progressed, exercise times became longer and so the 'spare' time between participant sessions was reduced. However, the instructor also felt that, as sessions got longer, it was important he guided participants through some stretches directly after the run to prevent any injury. This would sometimes result in the next participant arriving whilst the previous participant was still stretching. At that point, the instructor would welcome the next participant, get them set up with the heart rate monitor and on the treadmill, launch the robot-led session and return to finish off the previous participant's stretching whilst the next participant completed the (fully automated) warm up. As well as being a clear example of unplanned system use emerging through use, this is also an excellent demonstration of the potential such a system has to be useful in the real world. Further, this behaviour seems to have played a significant part in shaping how the instructor perceived the robot.

[KW]: Did you perceive the robot like a team-mate or a colleague or was it just a tool for you? How did you feel about Pepper?

[instructor]: *'It was definitely more of a colleague than a tool. Like Pepper wasn't a dumbbell or a kettlebell It was a colleague that you interact with in the gym. You know, I like to think her maybe early bugs or quirks definitely gave her a bit more of a personality that maybe I held on to. And obviously each of the clients would interact with her differently, which also gave her sort of more of a personality. And I could definitely feel, especially when the sessions got busier, runs got longer I had less time between each client being able to sort of work with and interact with Pepper'*

like you know, sort of dictating or coordinating other tasks while I was busy doing something else. So we could work together doing separate things but to get more work done and I think that's that's more of a teammate colleague trait than a tool. I'd like to see my hammer build my table for me.'

The instructor was further questioned regarding his role versus the robot's and how he thought participants might perceive them:

[KW]: When you were working with participants in a session. What do you think your role was versus the role of the robot in interacting and working with that person?

[instructor]: *'I feel, depending on what stage of the program really, I guess the main bulk of the program was me teaching sort of the AI what to say. I would say for the main majority of the study I was a teacher figure that would work with the robot and the robot would be the trainer that interacts with the client. And I sort of, I kind of just make comments on that robot's performance interacting. And obviously the client doesn't know that I'm sort of reinforcing or praising, that robot's sort of actions and behaviours. So it's quite. It keeps me separate from the client, but still in the loop of the whole thing.'*

[KW]: So do you feel like you're separate from the client? Do you think that that's how they perceived it as well? I mean, do you really feel like when they were working with the robot, they were working with the robot? Do you feel like you were very involved in that?

[instructor]: *'I think the majority of the time, it was the client and the robot. But there were there were always times where like the fourth wall was broken so to speak especially in in sort of scenarios that we couldn't predict like a shoe lace comes untied or someone needs a wee. And so they often would look to me rather than the robot as the instructor. So I feel that the lines were blurred but for the majority. It was. It was Pepper.'*

In the Phase 3 questionnaire, participants were explicitly asked to briefly describe the role of the robot, the fitness instructor and the researcher (author) in the context of delivering the programme. Their answers are presented in full in Appendix D. Participants appeared to identify separate but complimentary jobs performed/roles taken by the robot and the instructor, specifically that the robot delivered the exercise programme with the instructor being a background (but motivating) presence who also aided with stretches and therefore confidence in the programme. This reflects the instructor's comments above regarding participants working with the robot for the majority of the time. The instructor was also asked the potential impact of his presence during the post-study interview, and recognised (i) he likely did have some impact on participant motivation but also (ii) that his relationship development with participants was less than it would be in a traditional training scenario:

[instructor]: *'I think it's definitely interesting how such sort of complex or perceived complex social behaviours and work tasks can be boiled down to sort of simple instructions and behaviours'*

[KW]: Do you feel like you had a social influence on participants as well as the robot? Do you not think that they were partly there to see you and maybe even to some extent me, to please me

as the researcher as well as interacting with them?

[instructor]: *Definitely. Definitely. But I think at least having that kind of robot instructor still offers a standard like. Just knowing what it knows.'*

[KW]: But what do you think or to what extent do you think the participants still showed up and gave their effort for you even though they were working with the robot?

[instructor]: *'I think I definitely still had some underlying sort of effect or some sort of motivation because, you know, I did care about their, uh, their progress and their health. And so obviously, that will be a motivator. But whether, whether was just down to me, whether whether any instructor would still have that presence, I don't know.'*

[KW]: Do you feel like you made a relationship with the participants? Like a client relationship? You feel like you had a rapport with them?

[instructor]: *Yeah. To some degree, To some degree. Not as much as I would, let's say of the robot hadn't been there. Definitely. Definitely.*

Together, these results demonstrate the complexity of interactions and mutual shaping between the fitness instructor, robot and participants. The main exercise interactions were mostly between the participant and the robot, and teaching of the IML system was completely between the fitness instructor and the robot. However, it's clear that overall engagement with/delivery of the programme ultimately represents a triadic interaction between the instructor, participants and the robot. At odds with traditional lab-based HRI studies that often explicitly aim to reduce external human presence of e.g. the researcher or other bystanders, this represents a more holistic and realistic consideration of HRI in-the-wild as might actually result from real world robot deployment. Such consideration is both academically interesting, but also crucial to the development of effective systems that will have maximum positive impact on deployment.

An important element of mutual shaping *not* considered here is if/how/to what extent the suggestions made by the Learner of the IML system may have influenced the fitness instructor. For example, had the Learner *not* been making suggestions, such that the robot was entirely controlled/teleoperated by the instructor, would the action distribution and timing of actions remained the same? Further, if the instructor did not have the ability to actively reject suggestions (indicating that the Learner was not producing appropriate robot behaviour) would he still have identified those actions as being inappropriate? This is particularly interesting given the high number of suggested actions still being rejected at the end of Phase 2, immediately followed by seemingly appropriate robot behaviour overall positively evaluated by the instructor himself in Phase 3. Success of this approach inherently assumes that the domain expert/system 'teacher' would provide a 'correct' and fairly consistent response; i.e. that their unprompted actions and accept/reject decisions would not be overly affected by the type and timing of system suggestions. The results in Chapter 4 suggest that the IML approach does fundamentally 'work' for automating robot behaviour, but these questions are directly linked to its efficacy and thus warrant further investigation.

In addition, something difficult to measure or directly observe is how participants' themselves may have adapted to use/interaction with the robot, and the impact this may have had on the quality of interactions. This is specifically highlighted by GB's comment, highlighted in the previous subsection:

[GB]: *'It's hard to say whether the [IML robot] actually developed or whether my familiarity with it just improved, but it seemed like towards the end it's commenting was at the right frequency, whereas earlier on in the program I thought it was commenting too much or sometimes too little. I also don't know if the types of comments it chose changed, but I certainly found what it said to me more useful at the end than it did at the beginning.'*

Here, GB essentially describes an inability to objectively evaluate any changes in the robot due to recognising that their perception of/familiarity with the system was similarly dynamic over the course of the programme. Traditional lab based HRI studies will often consider this within the context of 'novelty effects' - i.e. that a robot may be evaluated overly-positively in early sessions with naive users due to the novelty of working with said robot. With repeated interactions, this effect would be expected to wear off, such that evaluations of the robot may fall and/or plateau. However, GB's comment points to a slightly different phenomenon, the potential for increased utility and effectiveness through user adaptation to/familiarity with the system. The relatively fixed nature of the experimental schedule and delivery of the couch to 5km programme reduced the potential for system use to significantly impact e.g. participant approach to exercise or exercise behaviour, however if a system like this were to be deployed in the home, these kinds of effects would likely be more significant, and so should also be considered in order to fully understand the impact of system deployment. Specifically, they might also offer some 'hope' against the general idea that once novelty effects have worn off, social robots would cease to be engaging and/or therefore useful.

5.4 Conclusion

This chapter explains how the generalisable methodologies employed for (i) the study with therapists of Chapter 2 and (ii) development of an autonomous SAR in Chapter 4 support a mutual shaping approach to SAR design and development. These methodologies are situated with respect to existing design methods for mutual shaping and how these have been applied in the context of SAR design, development and evaluation. Additional results from those works are also presented in order to demonstrate (i) why taking such an approach is worthwhile and (ii) evidence of mutual shaping that emerged during their implementation. To this end, key findings demonstrated in this chapter can be summarised as follows.

On focus groups for mutual shaping:

- Allowing broad discussions ahead of researchers presenting their proposed application(s)/research

agenda allows for identification of applications, use cases, opportunities and concerns that the researchers may not have considered/been aware of.

- A focus on mutual learning between researchers and domain experts allows for more informed discussion then relating to the proposed use case(s), with participants able to identify potential mutual shaping effects that would impact real world robot deployment and use/effectiveness.
- This focus on mutual learning has the potential to significantly improve participants' acceptance of robots.

On interactive machine learning (IML) as participatory design:

- As a method for designing and automating SAR behaviour, IML can result in a *better* (larger action space/more varied behaviour, evaluated more positively by users and a domain expert) autonomous system than can be achieved with a heuristic based approach.
- The IML approach also better supports a mutual shaping approach to robot design, as it allows for expert shaping of autonomous robot behaviour to reflect any desirable changes resulting from changes in the context of use/observation of mutual shaping effects on (long-term) deployment.
- The real world deployment of the robot utilised for training the IML system allows for the observation of mutual shaping effects that could not be captured in laboratory based studies (and as above, robot behaviour can be adapted reactively to such effects if required).

Combined with the results in Chapter 2 and 4, these findings demonstrate the synergy between wanting to pursuing a responsible approach to robot design/development as well as develop systems that are going to be effective in the real world. Across both methodologies, the focus on stakeholder inclusion and broad consideration of how robot deployment may impact on the social environment in which it is used, make them responsible design methodologies. For the IML process specifically, having an expert human-in-the-loop also reduces ethical risks regarding user safety and de-humanisation (discussed in Chapter 4). However, practically and technologically speaking, both methodologies were also demonstrated to be '*the right tool for the job*'. Results from the focus groups were sufficient to generate a rich set of design guidelines that informed later work in the thesis as well as representing a significant research contribution in their own right. Similarly, the IML process was demonstrated to result in a *better* autonomous system than the 'next best' participatory design option of heuristic based system design. As such, the work presented in this chapter should also provide strong motivation for pursuing a responsible, mutual shaping approach to robot design/development, rather than technologically

deterministic approaches that might consider ethical considerations and responsible innovation as a ‘checkbox exercise’ that may actually hinder research processes.

5.4.1 Limitations

A key limitation regarding how both of these methodologies were applied in this work is the lack of including system users (therapy patients and couch to 5km participants respectively) during the design/evaluation processes. As discussed in the introduction, inclusion of *all* stakeholders is a key aim of mutual shaping approaches to robot design/development. It is important to note that this is not a limitation within the methodologies themselves. The focus group methodology described in Section 5.2 would be equally well suited to exploring robot design with users as with domain experts. In its application to the study with therapists, the focus was simply on working with domain expert practitioners, in the first instance, because the purpose of the study was to inform the initial design of a robot that would play a similar role to that of those practitioners.

For the IML methodology, it is harder to imagine how e.g. an end user could be the expert-in-the-loop providing training examples and feedback, specifically in the context of SARs which essentially need to convince such users to undertake/stay engaged with a task they may not really want to do. Across both studies, a desire to include end users in the robot’s design and evaluation would raise the interesting issue of how users, who in theory are using the robot *precisely* because they themselves are not good at knowing/implementing what they need to stay motivated, can make the most useful input to these processes. For example, including users in initial detailed co-design of e.g. encouraging actions may not be appropriate, but they could certainly be included in preliminary testing of those actions designed with a domain expert. However, this also raises a number of interesting research questions regarding e.g. if, how and whether having one end user act as the expert-in-the-loop training the robot for *another* end user might actually impact on that end user’s own self-understanding and motivation to engage with the task. The impact that training a robot via the IML process might have on the human-in-the-loop providing that training data is another aspect of mutual shaping that could be considered in more detail in future works.

CONCLUSION

The overall goal of this work was to demonstrate a fully autonomous, socially assistive robot (SAR), developed using an expert-informed and mutual shaping approach, and deployed meaningfully in the real world. This was successfully achieved with the C25K robot coach described in Chapter 4, informed by the study with therapists in Chapter 2 and results regarding persuasive SAR behaviours in Chapter 3. Chapter 5 then identifies exactly *how* a mutual shaping approach was taken through the work as well as demonstrating *why* that proved valuable. As *individual* work packages, each of these chapters address a specific set of research questions to make one or more contributions to state of the art. A brief overview of each chapter and how the work progressed is given below.

Chapter 2 presented a qualitative study with therapists designed to understand (i) how they, as experts in socially assistive interaction, have an impact on service users' engagement with prescribed exercises and (ii) exactly what role a SAR might take in attempting to support that. A novel focus group methodology, with a focus on mutual learning, was designed to support a mutual shaping approach to this study. Chapter 5 presents results suggesting this methodology positively impacted on participants' acceptance of SARs, and generated significant additional insight into important considerations for real world SAR deployment. The results demonstrated that therapists knowingly use social interactions (purposefully tailoring their social behaviour) in order to affect changes in user behaviour through social influence. All of the findings were used to produce a generalisable set of design guidelines for SARs.

Informed by the study with therapists, Chapter 3 discussed whether task-focused, socially assistive HRI could be modelled as persuasion, based on social influence literature regarding human-human interaction (HHI). Specifically, it was posited that the Elaboration Likelihood Model (ELM) of persuasion from HHI might be able to usefully inform SAR design. A laboratory

based HRI study demonstrated that two out of three persuasive strategies derived from the ELM objectively increased the persuasiveness of a SAR in the context of encouraging exercise repetitions. This study also considered the acceptability of these persuasive behaviours and their purpose, in recognition of the fact that they might be considered *ethically hazardous* according British Standard BS 8611 on the ethical design of robotic devices (BSI 2016). The significant majority of participants found them to be acceptable (even if potentially deceptive) for the proposed use case.

Chapter 3 also presented another two preliminary, online, studies to investigate the potential impact of explicitly designing such behaviours in a way that better complied with the recommendations given in BS 8611. The results suggested that in some cases it might be possible to be more effective whilst also being more ethical, specifically in the context of referring to a human expert when citing expertise or referring to medical/technical specifics surrounding the prescribed exercises. However, for the more general socially supportive SAR behaviours considered by this thesis (e.g. affectively encouraging the user) then it might be beneficial for the robot to present itself (anthropomorphically) as an independent social agent. An additional related finding was that preferences regarding these social behaviours can vary significantly, with some users likely to prefer having zero social interaction with the robot at all.

Chapter 4 tackled the overall goal of the thesis to design, automate and evaluate a SAR in the real world. A ‘C25K robot coach’, co-designed with a fitness instructor, was developed to guide and encourage people through the NHS Couch to 5km programme. The role and behaviour design for the robot coach was further informed by the results from Chapters 2 and 3. The robot was designed to be automated through Interactive Machine Learning (IML), initially teleoperated with the fitness instructor observing the robot-user interaction and generating robot actions via a teaching interface, with the system learning from these teaching examples and eventually making its own action suggestions. An experimental study was conducted to train the robot then test its resulting autonomous behaviour whilst putting 10 participants through the Couch to 5km programme. A total of 151 sessions were spent training the system, which was then allowed to run 32 sessions autonomously (2-4 per participant). Results demonstrated that the IML system successfully learned *what* actions to do *for whom* but had mixed performance regarding *when* to do them; specifically suggesting actions too often.

A heuristic based version of the C25K robot roach was designed alongside the IML system, to represent an alternative expert-informed approach to generating autonomous behaviour. This robot was also deployed during the gym based experimental study as a within-subject manipulation, with participants asked to compare the two different versions of the robot coach, without knowing if/how they were programmed or designed differently. The results, demonstrated in Chapter 5 suggest the IML approach did result in an overall ‘better’ system, but also demonstrated again the potential for variation of preferences both within and between participants. Chapter 5 further evaluates both approaches as design processes that can support a mutual

shaping approach to SAR design.

Bringing these chapters together, as a whole, this work addresses the following, broader research questions, as set out in the Introduction.

6.1 Getting Human, Domain Expert Knowledge into SARs (RQ1)

How can human, domain expert knowledge, particularly regarding intuitive and experience-based social/emotional skills, be captured and utilised in the design and automation of a SAR?

The study with therapists built on previous participatory and user-centered design literature to demonstrate how focus groups and interviews with domain experts can generate design guidelines for informing SAR. Through the study methodologies employed, therapists were able to identify a number of ways they impact on exercise engagement/adherence, as well as ways in which a SAR might also support that. However, when it came to more detailed descriptions of complex behaviours, e.g. how they might decide what type of approach to take with different service users, it became clear that these sorts of decisions and behaviours were based on *intuition*. As such, whilst therapists were able to list some of the factors that might inform their behaviour and approach, and identify some of the ways they might tailor their behaviour, they couldn't describe any *definable rule set* for linking the two.

Concerning automation of behaviour then, this points towards trying to learn from expert behaviour 'in action'. Sussenbach et al. (2014) successfully used such an approach, conducting ethnographic observations of a human fitness instructor and using the coded results to inform design of a motivational interaction model. An alternative approach would be to apply machine learning to such coded observation data. However, there are two key issues with attempting to use coded observational data in this way. Firstly, the complexity of these socially assistive human-human interactions makes them exceedingly difficult to code. A small number of therapy sessions were informally observed by the author alongside the focus groups and interviews described in Chapter 2. Reflecting what was said in the focus groups and interviews, therapist behaviour was seen to significantly vary across service users, based on a huge number of factors unique to each individual. As such, it remains difficult to imagine how such data could be coded in a useful, generalisable way. Secondly, even if feasible, this approach would assume that the ultimate goal for a SAR is just to exactly replicate HHI behaviour. This seems like an unnecessary limitation on SAR design, and does not support any potential for adjustment of automated behaviour in response to mutual shaping effects.

The IML approach demonstrated by Senft et al. (2019) and employed in this work addresses both of those issues. It puts the expert *directly* in the automation loop such that they can provide training examples and feedback to the system in *realtime*. As such, their training examples/supervisory behaviour can be guided by the same intuition and social intelligence

that guides their own interaction behaviours. Design of the naive system requires the expert to identify only (i) what the robot should be able to do and (ii) what ‘input factors’ might be relevant for informing its behaviour; without having to give a detailed *why* to link those together. This reflects the type of information generated by the focus groups and interviews of the study with therapists. In addition, co-design of the IML robot’s action space, combined with the realtime training setup further allows the expert to essentially teach the robot what *it* should do, not necessarily what *they* would do. This was very much reflected in the fitness instructor’s approach to the design and use of the C25K robot coach discussed in Chapters 4 and 5, best summarised by this particular quote from the fitness instructor’s session notes presented in Appendix D:

*“Pepper’s suggestions might not be what *I* would say in that exact same situation, however it doesn’t mean that what was said or suggested was ‘wrong’”*

and in his description of Pepper as a colleague, with ‘her own’ personality (from Chapter 5):

“It was definitely more of a colleague than a tool... I like to think her maybe early bugs or quirks definitely gave her a bit more of a personality that maybe I held on to. And obviously each of the clients would interact with her differently, which also gave her sort of more of a personality.”

Together then, this work shows (i) how, at an early design/research stage qualitative studies with experts can identify promising research leads and shape a design/research programme and (ii) for design of a specific SAR, practical co-design sessions ahead of robot deployment, combined with an IML approach to automation on deployment, allow for expert knowledge to be utilised through the entire design, development and automation process.

6.2 On the Value of Taking Expert-Informed and Mutual Shaping Approaches (RQ2)

Does application of expert-informed and mutual shaping approaches approach result in SAR behaviours that are successfully able to improve user engagement with a task / programme? Where ‘success’ considers multiple contributory factors e.g. acceptability?

The study with therapists presented in Chapter 2 was vital in identifying the importance of social influence within the context of socially assistive HHI, therefore leading to the successful demonstration of how social influence and persuasion might usefully inform socially assistive HRI. Specifically taking a mutual shaping approach to this study further resulted in really significant insight regarding factors that need to be considered for real world SAR deployment to be successful, e.g. the extent to which the users’ social circle might *support* or *inhibit* their use of the robot. Given that these factors can only be observed on deployment of a SAR in-the-wild; this further supports mutual shaping approaches later in the design/development process that allow for the observation of these effects.

This was demonstrated in real world deployment of the C25K robot coach, which allowed for observation of how use of the robot shaped and was shaped by its social environment. To this end,

6.2. ON THE VALUE OF TAKING EXPERT-INFORMED AND MUTUAL SHAPING APPROACHES (RQ2)

Chapter 5 discusses interactions between the IML system, the fitness instructor and participants. Given this work was focused on SARs that would *compliment* rather than *replace* expert-led HHI, the instructor's interactions with and utilisation of the robot, and emergent behaviours around that, were of particular interest. The best example of this was the way in which, towards the end of the study when there was little free time between participants, the instructor started using the robot to warm-up the next participant whilst he continued to direct the current participant through stretches. The way he described this in a post-study interview (presented in Chapter 5 in full) further demonstrates how the approach taken successfully delivered on that initial motivation of wanting to compliment expert-led HHI:

"I could definitely feel, especially when the sessions got busier, runs got longer; I had less time between each client, being able to sort of work with and interact with Pepper like you know, sort of dictating or coordinating other tasks while I was busy doing something else. So we could work together doing separate things but to get more work done."

Participant acceptance of the C25K robot coach and engagement with the exercise programme seems also to have been positively impacted by them knowing there was an expert involved, with participants referring to the presence and role of the fitness instructor when discussing their willingness to work with the robot and motivations for continuing to attend exercise sessions and complete the programme. The way such experts can *directly* impact on acceptability is further discussed under RQ3. This RQ2 is more concerned with demonstrating whether these approaches would *inherently* lead to the development of SARs that were more acceptable (even if participants *didn't* know that such an approach had been taken). There are two key findings that support this notion. First is important insight regarding potential mutual shaping effects raised during the study with therapists. The therapists specifically highlighted many issues that need to be considered for broad stakeholder acceptance on real world deployment. These included issues that non-therapist robotocists would not necessarily have identified, e.g. relating to how SAR use might impact on or be impacted on by the user's broader social circle. Second is the acceptability of the SAR behaviours that were demonstrated in the persuasion studies of Chapter 3. These behaviours were designed by drawing from HHI literature identified based on the study with therapists results, as well as being informed by those results directly. When discussing acceptability of these behaviours, participants made a number of references to them being '*equivalent to what a human would do*'; further suggesting that SAR behaviours designed and based on acceptable, human expert behaviours would potentially '*inherit*' that acceptability.

More generally results regarding SAR acceptability and deception from the persuasion studies in Chapter 3 demonstrated how perception and acceptance of SAR behaviours is not independent from their *intention*, i.e. the overall context in which they are to be used. This would suggest it is likely *impossible* to assess acceptability and other perceptions of a SAR (e.g. credibility) without clearly establishing the context of use. Experimental HRI study design should therefore ensure participants are made aware of the proposed application, and utilise contextually valid

interaction scenarios (e.g. improving exercise engagement in this case) before asking them to complete any perception based measures (including e.g. the Godspeed questionnaire (Bartneck et al. 2009)). This further demonstrates the need for taking a mutual shaping approach that accounts for the overall context of proposed robot use.

Finally, the research philosophy provided in the Introduction described a desire to show how these approaches were not only necessary for acceptability, but were also *practically* and *technically* the ‘right tools for the job’. The co-design and IML approach used to create the C25K robot coach clearly demonstrates that this is the case. The difficulty in automating complex SAR behaviours was highlighted in the Introduction, with the only two previous examples of automated behaviour on SARs for exercise also using expert-informed approaches (Sussenbach et al. 2014, Martinez-Martin & Cazorla 2019). A number of results demonstrated by the C25K gym study were *only* possible because of the approach taken. For example, the robot C25K robot coach was shown to generate intelligently personalised behaviour across users when operating autonomously. This reflects how the fitness instructor personalised the robot’s behaviour across participants during the training phase; which in turn was only possible with the longitudinal, in-the-wild IML approach employed. Specifically, this approach allowed the instructor to observe and ‘get to know’ participants, and then to shape and tailor the robot’s behaviour accordingly.

6.3 The Role of Humans in Socially Assistive HRI (RQ3)

Considering a SAR within its broader context of use, what is the role of the human (i) designers/programmers behind its design/development and (ii) expert practitioners working with the robot ‘in-the-wild’?

The persuasion studies presented in Chapter 3 demonstrated two ways in which humans can have an impact on SAR credibility and acceptance. Firstly, participants of the lab-based exercise study referred to the idea that a SAR would be ‘signed off’, having been designed by relevant experts and then assessed and found to be safe, ethical etc, by the appropriate (human) authority. Secondly, results from the online ‘source of expertise’ study demonstrated that having a SAR explicitly reference an appropriate human authority, when discussing task related expertise, could have a positive impact on its acceptability and credibility. This seems to be in line with the concept of *inherited credibility* in persuasion; whereby credibility of a source can be improved if that source is introduced or endorsed by someone else who is perceived to be credible. Of course, it seems obvious that a therapy user is more likely to use a SAR if recommended to do so by their therapist. Possibly more nuanced however, the study with therapists also highlighted how other people around the user might influence use and acceptance of a SAR. This is another reason that a mutual shaping approach which accounts for this larger range of stakeholders, is so fundamental to successful SAR design.

The impact of *explicitly* having an expert involved in both the design *and* use of a SAR was

very well demonstrated by the C25K gym study. Participants explicitly referred to the fitness instructor's presence as giving them confidence to work out with the robot, as they knew a professional was on hand to ensure they were exercising safely. Further however, his presence had its own motivational impact, and specifically the results suggest that was *in addition* to, and not in any way *at the expense of* the impact of the robot itself. In discussing the role of the fitness instructor and the robot coach, participants were able to identify the separate, but complimentary roles taken by each, and how they both had a positive impact on their motivation.

This might be considered somewhat at odds with traditional HRI studies, in which a lot of effort is often taken to isolate the robot and participant *away* from the researchers or other humans e.g. through the Wizard-of-Oz paradigm (Riek 2012). This may well be necessary and important for some studies, e.g. when considering applications in which the robot will be working alone with users or, in this work, when investigating social influencing phenomena associated with the robot's presence (the persuasion studies in Chapter 3). However, these results show that, for SARs, having humans explicitly 'in the loop' during research, design and automation, *and* around their use on deployment, has a positive impact on the resulting outcomes. Arguably, this also better reflects how SARs are likely to be used in the real world, compared to those more traditional lab-based HRI studies that look to shield participants away from 'the person behind the robot'.

6.4 Ethical Considerations

This work was not focused on contributing to the fundamental ethical arguments surrounding SAR use. However, ethical implications were considered, practically, when designing and implementing the demonstrated SAR behaviours. Specifically, it was pointed out that social HRI behaviours are definitely somewhat at odds with recommendations made by a published standard on the ethical design of robotic devices (BSI 2016). Mostly, this is because many social HRI behaviours appear to actively encourage anthropomorphism, and imply social/emotional capabilities that the robot doesn't actually have. The work in Chapter 3 included some preliminary studies to investigate whether better compliance with this standard might have an impact on perception of the robot. The results suggested that in some cases, it might be possible to be more effective whilst also being more ethical, i.e. by referring to a human third party when it came to technical/medical expertise). In contrast, specifically for the kind of socially supportive social behaviours considered by this thesis, there is potentially good reason for the robot to present itself (anthropomorphically) as a social agent that appears to have concern for your welfare.

This was also found to be overwhelmingly acceptable to participants, and so may represent one use case for which such behaviours are justified. The standard does acknowledge that such use cases may exist, and identifies user validation as being one of the mechanisms for assessing related ethical risk. However, as discussed previously under RQ2, much of this acceptability was

linked with the specific application for which these behaviours were demonstrated (using a SAR to promote exercise engagement) and so may not hold for other applications or social HRI more generally. In addition, two key observations suggest it is important for future work to further consider whether these types of behaviours have the potential to be *exploited* in a more unethical way.

Firstly, participants appeared overwhelmingly undeterred by the potential for social robots to be deceptive, even when they were specifically primed to consider it (Studies 1 and 3 in Chapter 3). Participants felt very much that they knew the limitations of the robot, and were not in any way deceived as to its nature. However, the exercise based study in Chapter 3 demonstrated how manipulation of these behaviours could simultaneously have an *objective impact* on user behaviour. This suggests there might be a risk that such behaviours could be used to shape user behaviour *without them realising*. Given recent scandals regarding e.g. the use of social media to shape voting behaviour (Bond et al. 2012) this possibility is deserving of serious consideration in future work.

6.5 Limitations

A number of assumptions and constraints were identified when setting the scope of this work in the Introduction. These were deemed necessary to make the overall goal of the thesis, to demonstrate meaningful, real world deployment of a fully autonomous SAR, feasible in the context of a PhD research project. These included limiting investigations to one type and application of SAR, using a single, commercially available robot platform and focusing on the automation of *social* behaviour rather than e.g. implementing detailed kinematic assessment of user exercise performance.

Aside from these, the single biggest limitation of this work, given the focus on mutual shaping approaches, is the lack of diversity in stakeholder engagement during design and development work. The focus was very much on *expert* informed approaches, with little opportunity for *users* to be as equally involved. This was partly due to the use cases investigated, specifically whereby users typically enlist the help of experts because they themselves are struggling to get motivated, but further iteration e.g. of the C25K coach or development of a SAR for therapy would certainly warrant more user engagement in design, development and evaluation. This further lends itself to interesting avenues of future work; particularly regarding the role users might play in IML approaches to SAR automation. Very recent work has demonstrated that when it comes to personalising robot behaviour, it is better to have an *adaptive* robot, that autonomously tries to adjust to the users preference, rather than an *adaptable* robot whose behaviour the user can change directly (Schneider & Kummert 2020). This would suggest that it is better to have someone other than the user provide training data during IML, but leaves room to explore the impact of having a *peer*, e.g. a fellow therapy patient or exercise partner take on this role.

6.6 Future Work

This work used expert informed approaches to achieve real world deployment of an autonomous SAR. However, there are a number of variations that could be made to the approaches taken, as well as ways in which they could be extended to tackle additional complexity not considered in this work.

6.6.1 Variations on the Expert Role and Learning from the User

This work focused specifically on SARs used to take a somewhat *authoritative* role in instructing and encouraging the user through an activity, somewhat akin to the role an equivalent human authority (e.g. healthcare practitioner) would take. Accordingly, the term *expert*, in the context of the expert-informed approaches employed in this work, specifically referred to these human authority figures (i.e. the therapists in Chapter 2 and the fitness instructor in Chapter 4). However, key to a mutual shaping approach is the consideration of a *variety* of stakeholders, all of whom arguably represent some sort of *expert* with regards to system design. In the therapy use case for example, users are experts with regards to the barriers they might face in undertaking exercise at home, and so should be included in design work, even if overall behaviour design is primarily driven by the therapists' expertise. Including them in the *automation* of such robot behaviours is likely to be harder, given the authoritative role of the robot. Instead, user feedback might be incorporated *after* the SAR has been initially automated via SPARC (Senft et al. 2019), through reinforcement learning based on user input signals. This could be explicit (with the SAR asking the user for feedback) or implicit, based on the automatic analysis of user social signals (e.g. gaze). The latter seems more likely be effective, given the previously highlighted finding regarding *adaptive* versus *adaptable* robots (Schneider & Kummert 2020).

However, as identified in the Introduction, this authoritative role represents only one role a SAR might take in attempting to influence user behaviour. An alternative, particularly seen in other works considering SARs for children, is having the robot take the role of a *peer*, and potentially even requiring help such that the *user* actually represents an authority figure who then has a 'responsibility' to demonstrate the desired behaviours (Lemaignan et al. 2016, Cañamero & Lewis 2016). This opens up the possibility of having a peer, potentially someone who actually represents a potential user of the system, be the expert e.g. in the expert-in-the-loop, interactive machine learning (IML) approach used in this work. It may even be that *different experts* are used at different stages of the design and automation process. Consider the use of a SAR designed to aid children newly arrived to the country settle into school and make friends with the existing class population (Gillet & Leite 2020). It is reasonable to suggest that a teacher, play therapist or child psychologist might be an appropriate expert to co-design the SAR's action and input spaces and then automate the robot via IML. However, should the teaching interface be designed appropriately, it is also feasible that one of the *school children* could be invited to automate the

robot instead. A mutual shaping approach would then look to consider the impact this might have on *that child as well as* the children with whom the SAR interacts. For example, it would be interesting to whether teaching the SAR prosocial behaviours might also impact on *their own* prosociality.

6.6.2 Alternative Machine Learning Approaches to Utilise Expert Input

This work specifically employed SPARC (Senft et al. 2019) for automating socially assistive interaction behaviour for two key reasons, as outlined in Chapter 4. Firstly, it facilitates the learning of intuitive behaviour and tacit experience that is difficult to verbalise. Secondly, this learning can be done ‘in-the-wild’ and hence facilitate a mutual shaping approach to SAR automation. However, there are variations both within SPARC but also other machine learning options that could also meet one or both of these objectives with varying implications for the types of robot behaviours that can be automated.

Firstly, the specific KNN algorithm based setup used in this work means that the best possible outcome for the learning system is to exactly replicate the expert’s supervised use of the system; i.e. replicating exactly the same use of the action space. If, instead, the algorithm could learn to replicate and optimise the teacher’s *goal* rather than their *actions*; the resultant behaviours might actually be better than those demonstrated by the expert (Abbeel & Ng 2004). Even in attempting just to replicate the expert’s actions; the setup in this work was shown to fail in learning to sometimes *do nothing*; and so future works might consider how to improve on the specific setup employed here by better accounting for ‘do nothing’ as an action available to the learning system.

Another limitation of the demonstrated setup (and reflecting machine learning in robotics more generally) is that it was focused on one specific functional task, i.e. supporting the Couch to 5km programme for the robot fitness coach. As such, supporting another (even somewhat similar) task would require new design of the appropriate *task actions* (i.e. those relating to delivery of the prescribed activity). However, as described in Chapter 4 the overall learning framework was designed to be very generalisable. For example, considering the *social supporting actions* specifically (encouragements, jokes, asking how the user is feeling), the abstraction of specific actions to (*action-type, style*) pairings means those same pairings could be applied to other exercise based tasks (including cognitive rather than physical exercises). Similarly, whilst the input feature space included task specific measures (e.g. heart rate, speed) these also represented the more abstract features of task *engagement* and *performance* respectively, and hence could be replaced by alternative, equivalent measures for different tasks.

As such, if there was a way to easily implement new *task actions*; it might be possible to share the learning of *social supporting* actions from one activity to another within one SAR system to result in a multi-functional SAR that could guide users through a range of activities. Very recent work on Learning from Demonstration (LfD) proposed a framework that might support exactly

this. Louie & Nejat (2020) successfully demonstrated an LfD setup that allows non-robotocist care home practitioners to teach a SAR multi-step recreational activities, without the need for any robotocist involvement, that the SAR can then autonomously deliver to residents. However, this learning of the task is done *offline* i.e. ahead of deployment and therefore not including any robot-user interactions. This means there is (i) less potential to capture ‘in the moment’ intuitive behaviours and (ii) no ability to reflect mutual shaping effects related to robot deployment and users’ interactions with the robot.

Bringing this approach together with SPARC would seemingly offer the exciting potential to (i) have a domain expert teach the robot multiple activities (rather than having one single task functionality pre-decided and hard-coded before deployment), (ii) utilising the same machine learning model for *social supporting action* generation across multiple such activities and (iii) refining resultant SAR behaviours (initially taught *offline* as per Louie & Nejat (2020)) through expert-in-the-loop interactive machine learning and feedback (as per SPARC), therefore supporting the potential for mutual shaping.

6.7 Concluding Summary

This work set out to demonstrate the meaningful, real world deployment of a fully autonomous SAR developed using mutual shaping and expert-informed approaches. This was successfully achieved in the deployment of a robot ‘fitness coach’ deployed in a university gym to guide 10 participants through a National Health Service exercise programme. The work undertaken to achieve this goal resulted in a number of contributions to the field of socially assistive robotics as well as social HRI more generally, and also built on the very latest work in interactive machine learning for automating social HRI behaviours. The original contributions of this thesis (with chapter and relevant publication references) can be summarised as follows:

- A detailed set of **generalisable design guidelines for SARs** being used to prompt/encourage engagement with a prescribed task, resulting from a significant study with therapists.
(Chapter 2, Winkle et al. (2018))
- Application of **persuasion as a way to model socially assistive HRI**, with a series of studies to test whether persuasion literature can therefore inform SAR design. Results demonstrate that **socially persuasive human-human interaction strategies, if utilised on a SAR, can objectively increase the persuasive effectiveness of that SAR**, in this case, resulting in participants undertaking an increased number of exercise repetitions.
(Chapter 3, Winkle, Lemaignan, Caleb-Solly, Leonards, Turton & Bremner (2019))
- Practical consideration of how the above behaviours, as well as others demonstrated in existing social HRI literature are somewhat at odds with a published standard on the

ethical design of robots, and a preliminary studies on how **designing more ethical social HRI behaviours might impact on their effectiveness** as well as **participant acceptability of such behaviours**.

(Chapter 3)

- **Technical advancement of state of the art in interactive machine learning**, extending the recent development of Supervised, Progressively Autonomous Control (SPARC, Senft et al. (2019)) to (i) include user personality data in learning input, to allow for personalisation of robot behaviour and (ii) generate low-level robot ‘mood’ as well as robot actions. Successful deployment of a SAR, trained and automated using this approach, in a meaningful, longitudinal study ‘in-the-wild’.

(Chapter 4, to appear at Robotics: Science and Systems 2020)

- Practical demonstration of how **expert-informed and mutual shaping approaches** can be taken throughout the SAR design, development and evaluation process; including development of a **novel focus group methodology** to support this, and results to demonstrate **why such an approach is worth taking**.

(Chapter 5, Winkle, Caleb-Solly, Turton & Bremner (2019))

It is further hoped that the work presented in this thesis presents a positive case for what can be achieved in *interdisciplinary* research projects, demonstrating how *technical* work can be so much improved through combination with e.g. *qualitative* design studies and experimental measures, as well as the benefits of pursuing a responsible innovation approach.



APPENDIX A

This appendix contains the following resources from the study with therapists as referred to in Chapter 2:

- focus group topic guide
- acceptance measure administered before and after focus groups (the *before* questionnaire specifically presented here, showing additional images and a definition of social robots not repeated on the *after* questionnaire)
- interview topic guide
- full coding scheme as applied to focus group transcripts
- full coding scheme as applied to interview transcripts

Learning How to Help: Social Robots in Therapy (Part 1: Study with Therapists)

Focus Group Study

Before the Focus Group Session	Information sheet Consent form Demographics/therapist information collection Acceptance measure
On the Day [Approx 1 hr]	<p>Room Setup</p> <ul style="list-style-type: none"> • AV Equipment for presentation • Group seating • Audio recording equipment • Water / tea / coffee facilities <p>Welcome & Introduction [5 mins] <i>Welcome and explain that 'we are going to start the focus group by discussing some aspects of therapy that I am interested in. Later we will take a break I will talk a bit more about the work I am doing but I want to get your views on a few things before we talk too much about that'.</i></p> <p>Housekeeping:</p> <ul style="list-style-type: none"> • Collect signed consent forms • Name labels • Make participants aware when turning on audio equipment (and hence withdrawal issue)
	<p>Pre-Demo Discussion [30 mins]</p> <p>Topic Guide:</p> <ul style="list-style-type: none"> • Round the group introductions – name and area(s) worked in/typical service user(s) worked with • [Expert establishment] '<i>What are your main goals when working with service users?</i>' • [Robot images on screen] Use of robots in supporting a therapy program – '<i>What do you think about using robots to support a therapy program? How do you think that robots might be able to do that?</i>' • Self-practice as part of a therapy regime – '<i>Do you prescribe self-directed exercises/tasks for your service users to complete at home? What might these be? What is the importance of such exercises?</i>' • Reporting of self-practice – '<i>Do you ask service users to report back or keep a record of self-practice? Do you think this is accurate?</i>' • Engagement and motivation – '<i>How do you monitor service users' engagement? Do you often find yourself trying to motivate service users? How might you try to do that?</i>' • [Post it note exercise] Factors affecting compliance with self-practice – '<i>What kind of factors do you think affect service users' compliance with self-practice exercises? The literature suggests... (on screen) Use the post it notes to rank these, as well as come up with any additional factors you can think of.</i>'
	Project Presentation & Demo [20 mins]

	<p>[1] Exercise Based Pepper guides user through repetitions of a simple arm exercise.</p>	<p>[2] Task Based Pepper prompts user through a sequence based task e.g. making a cup of tea or preparing a microwave ready meal.</p>
<p>Post-Demo Discussion [10 mins]</p> <p>Topic Guide:</p> <ul style="list-style-type: none"> • Demo feedback – <i>'Firstly I would like to get your feedback on the demos – these are really my first attempt at what a robot coach might look like. What do you think? How would you have done it differently?'</i> • Revisit use of robots in supporting a therapy program – <i>'Now that you have seen the demonstrations, I'd like to discuss again your thoughts on using robots in therapy and how that might be beneficial. What would a robot aid look like, how could it help?'</i> • Useful data that could be collected by the robot for use by the therapist – <i>'Using a robot as well as other sensor systems it is possible to collect a huge range of data from the service user. Thinking about the measures you might use to monitor service user progress, what information is likely to be most useful to you?'</i> 		
<p>Thank you/Debrief</p> <p>Housekeeping:</p> <ul style="list-style-type: none"> • Acceptance measure • Interview scheduling/reminders(?) 		

Social Robot Questionnaire 1

Project Title: Learning How to Help: Social Robots in Therapy

Study: Study with Therapists

Name:

Social robots are those that can take part in social interactions with humans. They might exhibit human-social characteristics such as expressing and perceiving emotions, making conversation, establishing/maintaining social relationships, using natural cues (e.g. gaze, gesturing) and exhibiting a personality/character [1]. Social robots might be humanoid/resemble some human characteristics but this is not always the case. A range of social robots are shown below to demonstrate this.

[1] Fong, T., Nourbakhsh, I. and Dautenhahn, K., 2003. *A survey of socially interactive robots*. *Robotics and autonomous systems*, 42(3), pp.143-166.



Figure 1: A selection of social robots

Please indicate how much you agree or disagree with the following statements by ticking the appropriate boxes which are ordered from strongly disagree to strongly agree with neither agree nor disagree in the middle:

	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree
I feel apprehensive about the use of social robots in therapy					
Social robots are somewhat intimidating to me					
I think social robots might be somewhat intimidating to service users					
I think using social robots in therapy is a good idea					
I think using social robots would make therapy more engaging for the service user					
A social robot would be useful in supporting a therapy programme					
I think that use of a social robot could improve the positive outcomes/success of a therapy programme					

A social robot would be **most** useful (please choose one):

When I am working with service users

OR

When service users are working alone

Interview Topic Guide

Set-up	Reminder consent form already signed Audio recording equipment Whiteboard/Flipchart paper & pens DoH Segmentation Model information sheets
Introduction	<ul style="list-style-type: none"> • Reminder that already signed consent form • Have a copy of the information sheet here just in case • Reminder of recording and withdrawal rights as per information sheet
Introduce Service Users	The two service users you've got in mind – let's assign them a pseudonym each. Can you describe them to me?
Behaviour Differences <i>(Use whiteboard or pen and paper to write these down)</i>	<p>So thinking about those two service users, what would you say is similar about them? And what is different?</p> <p>When thinking about service users that you work with differently or take different approaches with, what made you choose these two in particular?</p> <p>So can you describe to be the differences in how you work with X and Y? Why do you take those different approaches?</p>
Behaviour Differences cont. & Knowledge of the Service User <i>(Use whiteboard or pen and paper to write down the categories and descriptors for each)</i>	<p>It seems therapists recognise that different people work in different ways and so I would like to explore whether we can sort of group people in some general categories based on that. The closest thing I could find in the literature was some research done by the Department of Health around motivation to lead a healthy lifestyle. You can see they came up with five categories and they came up with a persona and some key traits for each. (show example & explain)</p> <p>Do you think you could place X and Y into one of these categories?</p> <p>What do you think of using these to categorise therapy users? What categories would you have?</p> <p>What descriptors would you use?</p> <p>(after initial free discussion put up prompts from literature & focus groups)</p> <p>Building on that a little bit more, how long would you say it takes for you to get to know a service user and to figure out the best approach to take with them? What kind of things might you like to know about them in</p>

	<p>order to help with that? Do you use any particular tools or techniques for that?</p> <p>Finally, thinking about each of these categories, and back to what we discussed about your different approaches with X and Y, could you give any particular approaches or ideas of what might work better for people in these categories?</p>
Feedback	<p>(Something about having a couple of more general questions)</p> <p>How do you utilise feedback during a therapy session?</p> <p>What might trigger you to give some feedback?</p>
Progress	<p>Focus group discussions have suggested that motivation might be linked to the service users' progress and perhaps reminding them of that. I've also heard a lot about goal setting, so I'd just like to ask how you approach something like that with somebody who has a long term condition or progressive illness?</p>

Focus Group Coding Scheme

First Level Coding Scheme. Those marked* have second level codes presented below:

Code	Description
[A] Application of Robots in Therapy	Direct suggestions or inferred applications/functionalities of SARs in therapy (<i>in post-demo discussion</i>)
[A] Therapist Opinion of Social Robots	Any feedback on social robots or their use in therapy – could be appearance, functionality etc. from the therapists' point of view (<i>in post-demo discussion</i>)
[A] User Opinion of Social Robots	As above but with therapist referencing user response
[B] Application of Robots in Therapy	Direct suggestions or inferred applications/functionalities of SARs in therapy (<i>in pre-demo discussion</i>)
[B] Therapist Opinion of Social Robots	Any feedback on social robots or their use in therapy – could be appearance, functionality etc. from the therapists' point of view (<i>in pre-demo discussion</i>)
[B] User Opinion of Social Robots	As above but with therapist referencing user response
Demo Feedback	Feedback on any aspect of the robot demonstrations
Factors relating to engagement	Anything about the factors which impact on adherence, including which factors are most important (e.g. around the ranking exercise)
Importance of self-practice	Any reflections on how important such exercises are to the overall therapy programme
Measuring motivation or engagement	Measures or indicators of how motivated/engaged a service user is
Mutual/ Social Shaping	Comments regarding societal influences in therapy, or issues likely to impact on real-world SAR deployment in therapy
Personalised Approaches	Comments regarding how therapy delivery or therapist approach might be adapted or personalised based on the service user
Prescription of Self-Practice	Anything describing the prescription of or detail about the type of self-practice exercises given to service users
Reporting of Self-Practice	Anything about whether the therapist monitors self-practice, or ways in which it might be measured
Robot Requirements	Anything around perceived difficulties the robot might face or useful references to what sensor/data collection capabilities might be required
Therapist Behaviour or Role	Anything about how the therapist might have an impact on motivation or engagement i.e. through taking a particular approach or adjusting their behaviour

Second Level Coding Nodes. Those marked^ have third level codes presented below:

First Level/ 'Parent' Code	Second Level/ 'Child' Codes
[A/B] Application of Robots in Therapy	Calming or anxiety Demonstrating or showing task

	Engagement in therapy^ Interpretation and Translation Medication Tele-operation Therapist or other feedback^ Therapist or other training User feedback
[A/B] Therapist Opinion of Social Robots	Positive Negative
[A/B] User Opinion of Social Robots	Positive Negative
Factors relating to engagement	Cognition Demographics Dynamic & individual Ease of access Enjoyment External feedback / encouragement / information^ Memory Mental health Mood & emotional state Routine Self-efficacy & expectations & ownership Severity of ailment Social^
Therapist Behaviour or Role	Boosting intrinsic motivation^ Feedback^ Make sessions enjoyable Observe & react to patient Persuasion & social influence / social interaction^ Scheduling

Third Level Coding Nodes:

First Level/ 'Parent' Code	
Second Level/ 'Child' Code	Third Level/ 'Grandchild' Codes
[A/B] Application of Robots in Therapy	
Engagement in therapy	Improving task value Prompting and facilitating Robot as third party Robot influence
Therapist or other feedback	Accurate reporting Quality/ ability of patient activity Useful data
Factors relating to engagement	
External feedback / encouragement / information	Progress based Effort based/ encouragement Functional/ medical purpose/ understanding
Social	Social influence/ pressure Social support
Therapist Behaviour or Role	

Boosting intrinsic motivation	Frame in term of functional goals Functional/ intrinsic motivators Improve patient understanding
Feedback	Effort based Feedback based
Persuasion & social influence / social interaction	'Cajoling' & accountability Expert support Interaction

Interview Coding Scheme

First Level Nodes	Second Level Nodes (if any)
Motivation strategy	
Personalised approaches	
Engagement factors	Functional goal Expectations Financial issues Social support
Description of the user	
Giving feedback	
Thoughts on NHS framework	
Instantaneous data/patient indicators	Self-assessment Performance; effort Bodily cues Pain; fatigue Symptoms Vitals Timing Enjoyment
Social shaping	
Thoughts on general categorisation	
Link to robot	
Measuring motivation/engagement	
Tools for knowing patient and personalisation	
Progress markers	
Excuses for low engagement	
Possible robot actions/functionalities	
Considerations for categorisation	



APPENDIX B

This appendix contains the following resources from the persuasion studies as referred to in Chapter 3:

- study 1 experimental measures in full
- study 2 experimental measures in full

Study 1 Experimental Measures in Full

After each interaction

Perceived Credibility [1]:

Please rate your impression of the robot on these scales:

Expertise

Unexperienced	1	2	3	4	5	Experienced
Uninformed	1	2	3	4	5	Informed
Untrained	1	2	3	4	5	Trained
Unqualified	1	2	3	4	5	Qualified
Unskilled	1	2	3	4	5	Skilled
Unintelligent	1	2	3	4	5	Intelligent
Incompetent	1	2	3	4	5	Incompetent
Stupid	1	2	3	4	5	Bright

Trustworthiness

Dishonest	1	2	3	4	5	Honest
Untrustworthy	1	2	3	4	5	Trustworthy
Close-minded	1	2	3	4	5	Open-minded
Unjust	1	2	3	4	5	Just
Unfair	1	2	3	4	5	Fair
Selfish	1	2	3	4	5	Unselfish
Immoral	1	2	3	4	5	Moral
Unethical	1	2	3	4	5	Ethical
Phony	1	2	3	4	5	Genuine

Goodwill

Doesn't care about me	1	2	3	4	5	Cares about me
Doesn't have my interests at heart	1	2	3	4	5	Has my interests at heart
Self-centred	1	2	3	4	5	Not self-centred
Not concerned with me	1	2	3	4	5	Concerned with me
Insensitive	1	2	3	4	5	Sensitive
Not understanding	1	2	3	4	5	Understanding

Extroversion

Timid	1	2	3	4	5	Bold
Quiet	1	2	3	4	5	Verbal
Meek	1	2	3	4	5	Aggressive
Silent	1	2	3	4	5	Talkative

Composure

Poised	1	2	3	4	5	Nervous
Relaxed	1	2	3	4	5	Tense
Calm	1	2	3	4	5	Anxious
Composed	1	2	3	4	5	Excitable

Sociability

Good-natured	1	2	3	4	5	Irritable
Cheerful	1	2	3	4	5	Gloomy
Friendly	1	2	3	4	5	Unfriendly

Perceptions of the robot [2]:

Please rate your impression of the robot on these scales:

Anthropomorphism:

Fake	1	2	3	4	5	Natural
Machinelike	1	2	3	4	5	Humanlike
Unconscious	1	2	3	4	5	Conscious
Artificial	1	2	3	4	5	Lifelike
Moving rigidly	1	2	3	4	5	Moving elegantly

Animation:

Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Inert	1	2	3	4	5	Interactive

Likeability:

Dislike	1	2	3	4	5	Like
Unfriendly	1	2	3	4	5	Friendly
Unkind	1	2	3	4	5	Kind
Unpleasant	1	2	3	4	5	Pleasant
Awful	1	2	3	4	5	Nice

Perceived Safety:

Please rate your emotional state on these scales:

Anxious	1	2	3	4	5	Relaxed
Calm	1	2	3	4	5	Agitated
Quiescent	1	2	3	4	5	Surprised

Questions:

- To what extent do you feel that you developed a relationship with the robot? (not at all / not much / not sure / a bit / a lot) [3]
- To what extent do you feel that the robot developed a relationship with you? (not at all / not much / not sure / a bit / a lot) [3]

Imagine you were undergoing a therapy regime where you had to do exercises like this every day, and you had this robot at home to help you in-between visits from your therapist:

- On a scale of 1 (low) to 5 (high) how much responsibility do you think your therapist would hold for monitoring your engagement with your exercises?
- On a scale of 1 (low) to 5 (high) how much responsibility do you think this robot would hold for monitoring your engagement with your exercises?
- On a scale of 1 (low) to 5 (high) how much responsibility do you think your therapist would hold for giving you advice about your symptoms and the exercises you do at home?
- On a scale of 1 (low) to 5 (high) how much responsibility do you think this robot would hold for giving you advice about your symptoms and the exercises you do at home?

Additional questions after seeing both robots

- Which robot do you think was the most motivating, and why? (robot 1 / robot 2)
- Which robot would you prefer to work with and why? (robot 1 / robot 2)

In the goodwill / similarity conditions:

- The main difference between the two robots you saw was that robot (1/2) was programmed to demonstrate (more goodwill / more expertise/ some similarity with you). There is some concern that these and other human-like social behaviours may be deceptive. For example, such robots do not and cannot feel emotions, nor do they have any real interest in the person they are interacting with. Do you feel that either of the robots you saw today were deceptive? (if so) Why? Would you be happy for the robot to act in the ways you saw today?

[1] R. H. Gass and J. S. Seiter, *Persuasion: Social Influence and Compliance Gaining*. Routledge, 2015.

[2] C. Bartneck, D. Kulic, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[3] J. Hall, T. Tritton, A. Rowe, A. Pipe, C. Melhuish, and U. Leonards, 'Perception of own and robot engagement in human–robot interactions and their dependence on robotics knowledge', *Robotics and Autonomous Systems*, vol. 62, no. 3, pp. 392–399, Mar. 2014.

Study 2 and 3 Experimental Measures in Full

After each video

Perceived Credibility [1]:

Please rate your impression of the robot on these scales:

Expertise

Unexperienced	1	2	3	4	5	Experienced
Uninformed	1	2	3	4	5	Informed
Untrained	1	2	3	4	5	Trained
Unqualified	1	2	3	4	5	Qualified
Unskilled	1	2	3	4	5	Skilled
Unintelligent	1	2	3	4	5	Intelligent
Incompetent	1	2	3	4	5	Incompetent
Stupid	1	2	3	4	5	Bright

Trustworthiness

Dishonest	1	2	3	4	5	Honest
Untrustworthy	1	2	3	4	5	Trustworthy
Close-minded	1	2	3	4	5	Open-minded
Unjust	1	2	3	4	5	Just
Unfair	1	2	3	4	5	Fair
Selfish	1	2	3	4	5	Unselfish
Immoral	1	2	3	4	5	Moral
Unethical	1	2	3	4	5	Ethical
Phony	1	2	3	4	5	Genuine

Goodwill

Doesn't care about me	1	2	3	4	5	Cares about me
Doesn't have my interests at heart	1	2	3	4	5	Has my interests at heart
Self-centred	1	2	3	4	5	Not self-centred
Not concerned with me	1	2	3	4	5	Concerned with me
Insensitive	1	2	3	4	5	Sensitive
Not understanding	1	2	3	4	5	Understanding

Extroversion

Timid	1	2	3	4	5	Bold
Quiet	1	2	3	4	5	Verbal
Meek	1	2	3	4	5	Aggressive
Silent	1	2	3	4	5	Talkative

Composure

Poised	1	2	3	4	5	Nervous
Relaxed	1	2	3	4	5	Tense
Calm	1	2	3	4	5	Anxious
Composed	1	2	3	4	5	Excitable

Sociability

Good-natured	1	2	3	4	5	Irritable
Cheerful	1	2	3	4	5	Gloomy
Friendly	1	2	3	4	5	Unfriendly

Perceptions of the robot [2]:

Please rate your impression of the robot on these scales:

Anthropomorphism:

Fake	1	2	3	4	5	Natural
Machinelike	1	2	3	4	5	Humanlike
Unconscious	1	2	3	4	5	Conscious
Artificial	1	2	3	4	5	Lifelike
Moving rigidly	1	2	3	4	5	Moving elegantly

Animation:

Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Inert	1	2	3	4	5	Interactive

Likeability:

Dislike	1	2	3	4	5	Like
Unfriendly	1	2	3	4	5	Friendly
Unkind	1	2	3	4	5	Kind
Unpleasant	1	2	3	4	5	Pleasant
Awful	1	2	3	4	5	Nice

Perceived Intelligence:

Incompetent	1	2	3	4	5	Competent
Ignorant	1	2	3	4	5	Knowledgeable
Irresponsible	1	2	3	4	5	Responsible
Unintelligent	1	2	3	4	5	Intelligent
Foolish	1	2	3	4	5	Sensible

Perceived Safety:

Please rate your emotional state on these scales:

Anxious	1	2	3	4	5	Relaxed
Calm	1	2	3	4	5	Agitated
Quiescent	1	2	3	4	5	Surprised

Questions:

- To what extent do you feel the patient developed a relationship with the robot? (not at all / not much / not sure / a bit / a lot) (adapted from [3])
- To what extent do you feel the robot developed a relationship with the patient? (not at all / not much / not sure / a bit / a lot) (adapted from [3])
- Who do you think is responsible for monitoring [actor name]’s engagement with her home exercises? (the robot / [actor name] / the therapist / other + blank)
- Who do you think is responsible for giving [actor name] advice about the exercises she does at home? (the robot / the therapist / other + blank)
- How would you describe the role and responsibilities of the robot shown in the video? (free space)

Once all videos have been watched

- Which robot do you think was the most motivating, and why? (free space)
- Which robot would you prefer to work with and why? (free space)
- Repeat of negative attitude to robots scale as administered in pre-study demographics and presented in ‘Demographics’

In Study 1 (Expertise) only:

- The robot told [actor name] lots of medical information about her injury, where did that information come from? (The robot / [actor name]’s therapist / the people who designed and programmed the robot / other + free space)

In Study 2 (Sociability) only:

- Having watched all of the videos, do you think any of the robots were deceptive? If so, please give details on which robot(s) and why. (free space)

After debrief

- Having read the debrief, do you have any more thoughts or comments you would like to add? Do you think any of your previous answers would change now, based on the debrief information? (free space)

[1] R. H. Gass and J. S. Seiter, Persuasion: Social Influence and Compliance Gaining. *Routledge*, 2015.

[2] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, “Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots,” *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[3] J. Hall, T. Tritton, A. Rowe, A. Pipe, C. Melhuish, and U. Leonards, ‘Perception of own and robot engagement in human–robot interactions and their dependence on robotics knowledge’, *Robotics and Autonomous Systems*, vol. 62, no. 3, pp. 392–399, Mar. 2014.



APPENDIX C

This appendix contains the following resources from the study with therapists as referred to in Chapter 4:

- information sheet included with advertisements for the Couch to 5km gym study
- post-session participant questionnaire
- weekly participant journal
- within-subject questionnaire implemented at the end of Phase 1
- within-subject questionnaire implemented at the end of the study
- examples of notes taken by the fitness instructor during the exercise sessions

Study Information Sheet

Study Title: Training a Robot Coach

Date: Summer 2019

Contact Address: Katie Winkle, Bristol Robotics Laboratory, University of the West of England, Coldharbour Lane, Bristol, BS16 1QY.

Email: k.winkle@bristol.ac.uk

Thank you for taking the time to consider participating in my research project. This information sheet gives an overview of the proposed work. If you decide to take part, you will give more information on exactly how the study will run.

Who is doing the work?

Katie Winkle, PhD student based at the Bristol Robotics Laboratory, University of the West of England.

What is the project and study for?

This study is part of PhD research aiming to design a social robot to function as a coach for guiding therapeutic exercises. The aim of this study is to have a human expert (e.g. physiotherapist or personal trainer) train the robot using supervised machine learning. Further, we wish to explore whether such an approach (and the resulting robot system) might be a useful tool – i.e. how people feel about, and their experience of, working with such a robot throughout a real-world, long-term exercise programme.

Who are we looking for?

We invite participants who meet the following criteria:

- 18 years old or over
- No health conditions preventing safe engagement with the NHS Couch to 5K programme
- Fluent in English
- Not currently running often

We are looking to recruit people who would like to take up running as a form of exercise – and would be interested in following an NHS-designed running plan, ‘Couch to 5K’ to do so. Your current fitness or running experience is not important so long as you do not frequently run already and are in general good health with no medical conditions which might prevent you safely taking part in the programme. The following guidance is taken from the NHS Couch to 5K website:

Who is Couch to 5K for?

Couch to 5K is for everyone. Whether you've never run before or if you just want to get more active, Couch to 5K is a free and easy way of getting fitter and healthier. If you have any health concerns about beginning an exercise regime like Couch to 5K, make an appointment to see your GP and discuss it with them first.

What does the study involve?

You will be invited to complete 9 week the Couch to 5K programme, attending 3x weekly exercise sessions at Wallscourt Farm Gym, Frenchay Campus. During these sessions you will be accompanied by the social robot Pepper and supervised by a qualified fitness instructor. The role of the fitness instructor will be to observe your interaction with the robot and generate training data for improving its behaviour. You are free to miss, rearrange or stop attending exercise sessions at any

time during the 9 week programme. Additionally you can stop or take a break at any time during an exercise session.

During the first few weeks of the study, you will have the chance to test out a few different versions of the robot. Throughout the study, we will ask you to complete a brief weekly journal documenting your experiences of undertaking the programme and working with the robot.

Before starting the study you will have the chance to meet the fitness instructor and go through the proposed exercise programme in detail. The following information is taken from the Couch to 5K website. Please note that whilst Couch to 5K is an NHS designed programme, this research is not being carried out in direct collaboration with the NHS.

What is Couch to 5K?

Couch to 5K is a running plan for absolute beginners. It was developed by a new runner, Josh Clark, who wanted to help his 50-something mum get off the couch and start running, too. The plan involves 3 runs a week, with a day of rest in between, and a different schedule for each of the 9 weeks.

How does Couch to 5K work?

Probably the biggest challenge a new runner faces is not knowing how or where to start. Often when trying to get into exercise, we can overdo it, feel defeated and give up when we're just getting started. Couch to 5K works because it starts with a mix of running and walking to gradually build up your fitness and stamina. Week 1 involves running for just a minute at a time, creating realistic expectations and making the challenge feel achievable right from the start.

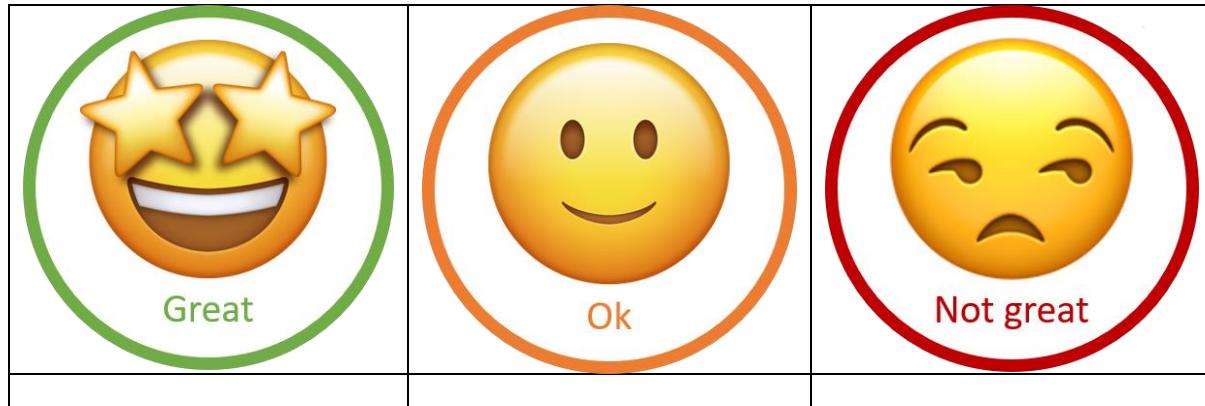
Contacts

If you have any questions about the topic of this research, or taking part in this study then please contact the researcher Katie Winkle (k.winkle@bristol.ac.uk) or project supervisor Paul Bremner (paul.bremner@brl.ac.uk).

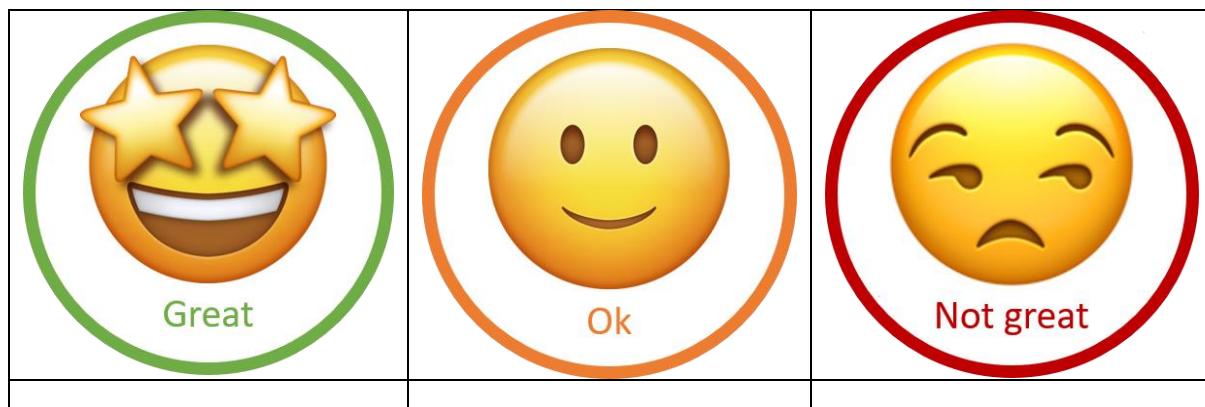
Participant Post-Session Questionnaire

UserID:	Session:	Robot Version:
---------	----------	----------------

How did you find today's session? (Please tick the appropriate box)



How would you rate the robot as an exercise instructor based on today's session?



Any thoughts or comments?

'Participant Weekly Journal'

User ID:	Week:
----------	-------

Please reflect on your experience of the Couch to 5km and working with the robot/different robot versions this week. How do you feel about the exercise programme? What are your thoughts on the robot as an exercise instructor and companion?

Intro

Over the last eight sessions you have worked out with two different versions of Pepper - Robot A (Orange) and Robot B (Purple). This questionnaire is designed to explore any differences in your perception of these robots and your experiences of working with them.

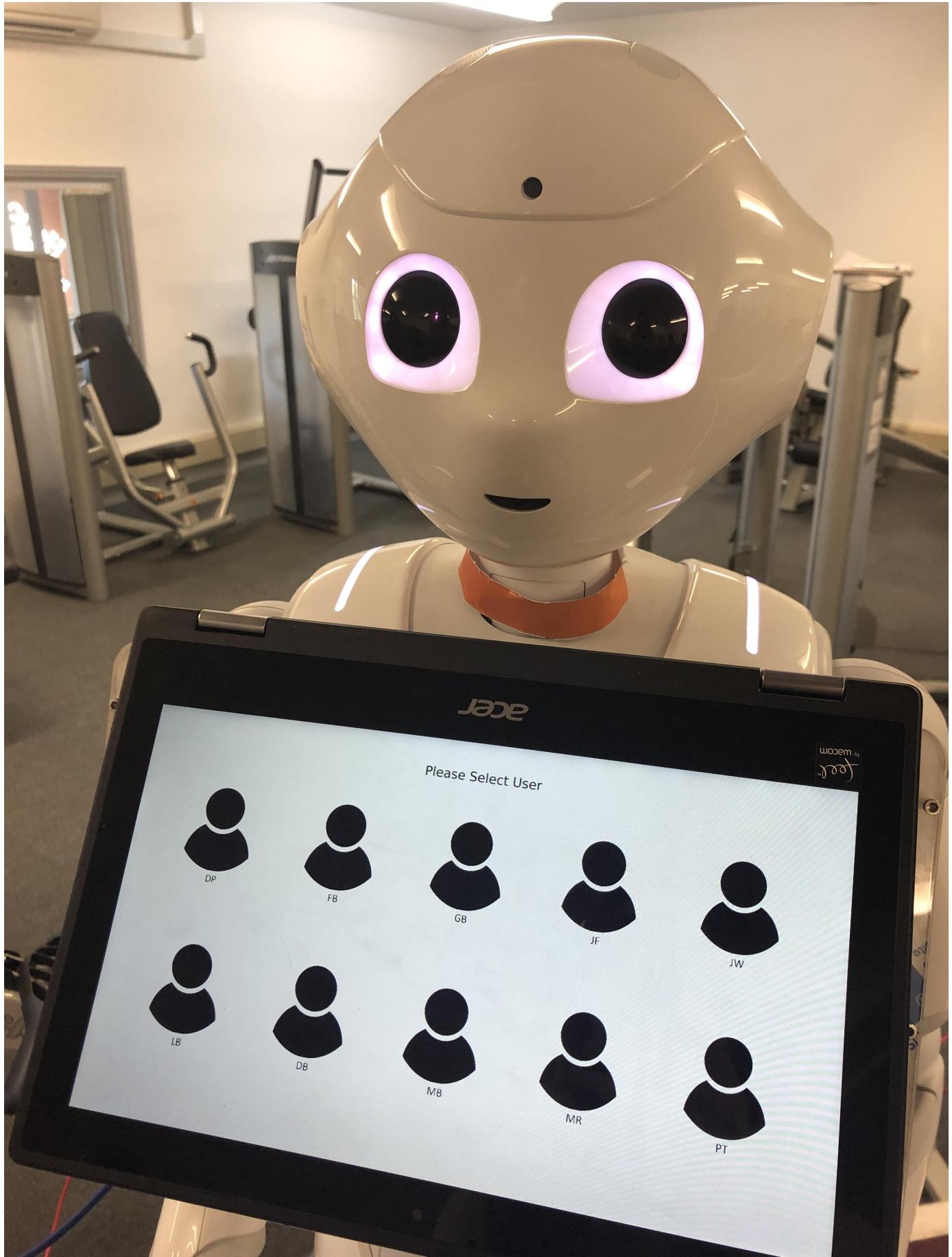
Block 4

Please enter your programme User ID (your initials):

Robot A: Orange

Please think about your experiences with Robot A (Orange) when answering these questions:

Robot A
Orange



As an exercise instructor, did you find Robot A (Orange) to be:

Unexperienced



Experienced

Uninformed	Informed
Untrained	<input type="radio"/>
Unqualified	<input type="radio"/>
Unskilled	<input type="radio"/>
Unintelligent	<input type="radio"/>
Incompetent	<input type="radio"/>
Stupid	<input type="radio"/>
	Trained
	Qualified
	Skilled
	Intelligent
	Competent
	Bright

Did you feel that Robot A (Orange):

Didn't care about me	<input type="radio"/>	Cared about me				
Didn't have my interests at heart	<input type="radio"/>	Had my interests at heart				
Was self-centred	<input type="radio"/>	Was not self-centred				
Was not concerned with me	<input type="radio"/>	Was concerned with me				
Was insensitive	<input type="radio"/>	Was sensitive				
Was not understanding	<input type="radio"/>	Was understanding				

Did you find Robot A (Orange) to be:

Irritable	<input type="radio"/>	Good-natured				
Gloomy	<input type="radio"/>	Cheerful				
Unfriendly	<input type="radio"/>	Friendly				

Please rate your impression of Robot A (Orange) on the following scales:

Dislike	<input type="radio"/>	Like				
Unfriendly	<input type="radio"/>	Friendly				
Unkind	<input type="radio"/>	Kind				
Unpleasant	<input type="radio"/>	Pleasant				
Awful	<input type="radio"/>	Nice				

Please indicate your response to the following questions:

	Not at all	Not much	Not sure	A bit	A lot
To what extent do you feel you developed a relationship with Robot A (Orange)?	<input type="radio"/>				

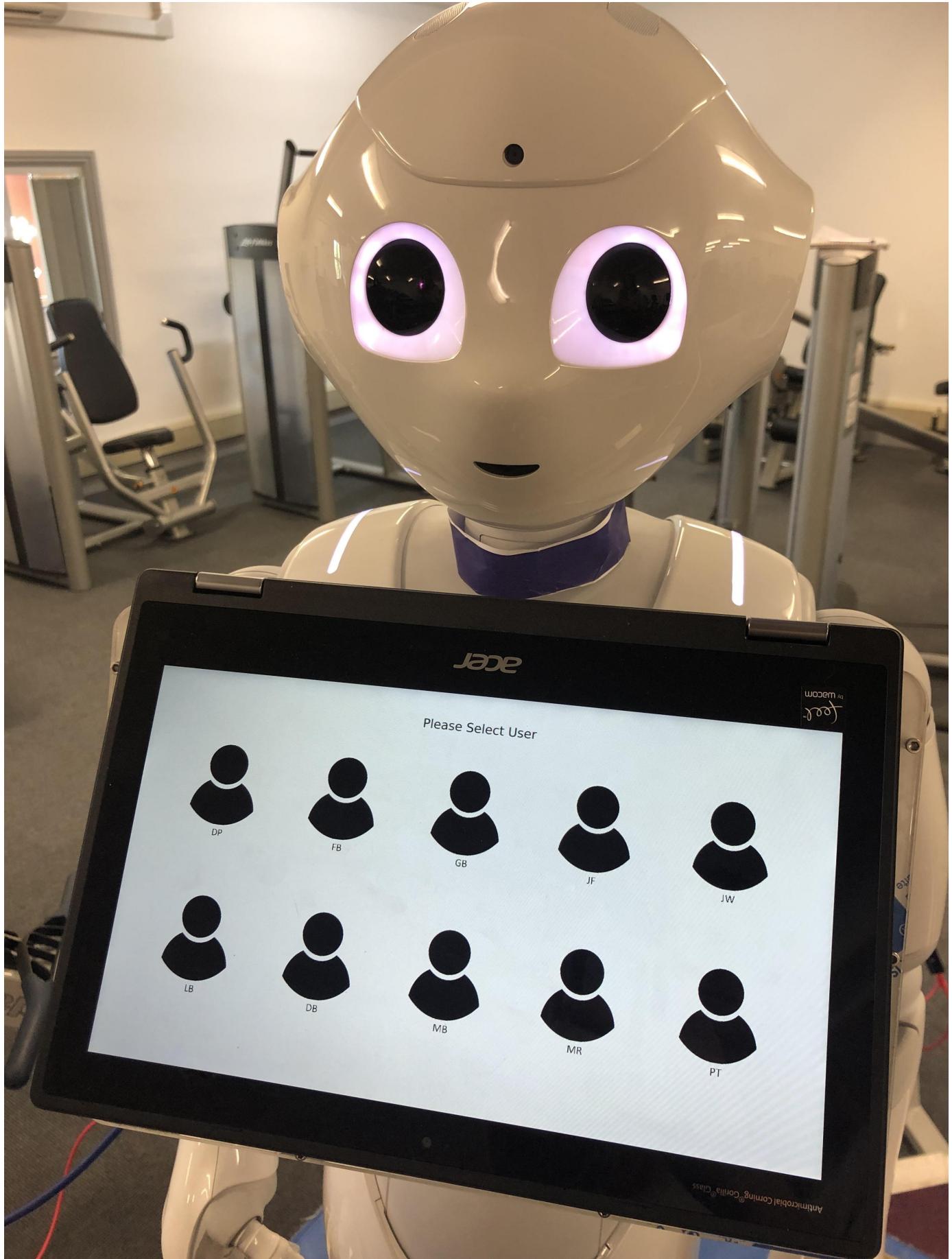
Not at all Not much Not sure A bit A lot

To what extent do you feel Robot A (Orange) developed a relationship with you?

Robot B: Purple

Please think about your experiences with Robot B (Purple) when answering these questions:

Robot B
Purple



As an exercise instructor, did you find Robot B (Purple) to be:

Unexperienced



Experienced

Uninformed	Informed
Untrained	<input type="radio"/>
Unqualified	<input type="radio"/>
Unskilled	<input type="radio"/>
Unintelligent	<input type="radio"/>
Incompetent	<input type="radio"/>
Stupid	<input type="radio"/>
	Trained
	Qualified
	Skilled
	Intelligent
	Competent
	Bright

Did you feel that Robot B (Purple):

Didn't care about me	<input type="radio"/>	Cared about me				
Didn't have my interests at heart	<input type="radio"/>	Had my interests at heart				
Was self-centred	<input type="radio"/>	Was not self-centred				
Was not concerned with me	<input type="radio"/>	Was concerned with me				
Was insensitive	<input type="radio"/>	Was sensitive				
Was not understanding	<input type="radio"/>	Was understanding				

Did you find Robot B (Purple) to be:

Irritable	<input type="radio"/>	Good-natured				
Gloomy	<input type="radio"/>	Cheerful				
Unfriendly	<input type="radio"/>	Friendly				

Please rate your impression of Robot B (Purple) on the following scales:

Dislike	<input type="radio"/>	Like				
Unfriendly	<input type="radio"/>	Friendly				
Unkind	<input type="radio"/>	Kind				
Unpleasant	<input type="radio"/>	Pleasant				
Awful	<input type="radio"/>	Nice				

Please indicate your response to the following questions:

	Not at all	Not much	Not sure	A bit	A lot
To what extent do you feel you developed a relationship with Robot B (Purple)?	<input type="radio"/>				

Not at all Not much Not sure A bit A lot

To what extent do you feel Robot B (Purple) developed a relationship with you?

Robot A: Orange versus Robot B: Purple

Now thinking about comparing the two robots:

How would you describe Robot A (Orange) and Robot B (Purple) as exercise instructors? What differences, if any, did you perceive between them?

[Redacted]

Do you feel that one of the robots encouraged you to perform better than the other?

- Robot A (Orange)
 Robot B (Purple)
 No preference

Did you prefer working with one robot over the other?

- Robot A (Orange)
 Robot B (Purple)
 No preference

If you could choose to work with one robot or the other for the remainder of the programme, would you have a preference? Please explain your reasoning.

- Robot A (Orange)
 Robot B (Purple)
 No preference

Please feel free (but not obliged) to leave any additional thoughts or comments here:

Powered by Qualtrics

Intro

In this study we've been working with two different versions of Pepper. Half of our participants (including yourself) have been seeing the Purple robot, whereas the other half have been seeing the Orange one. Now that we're near the end of the study we wanted to show everyone both robots again for a final comparison. You should have just completed a session with the Orange robot. This questionnaire is designed to explore any differences in your perception of the two robots and your experiences of working with them. In addition, there are some final questions about your overall experience of taking part in this study.

Block 4

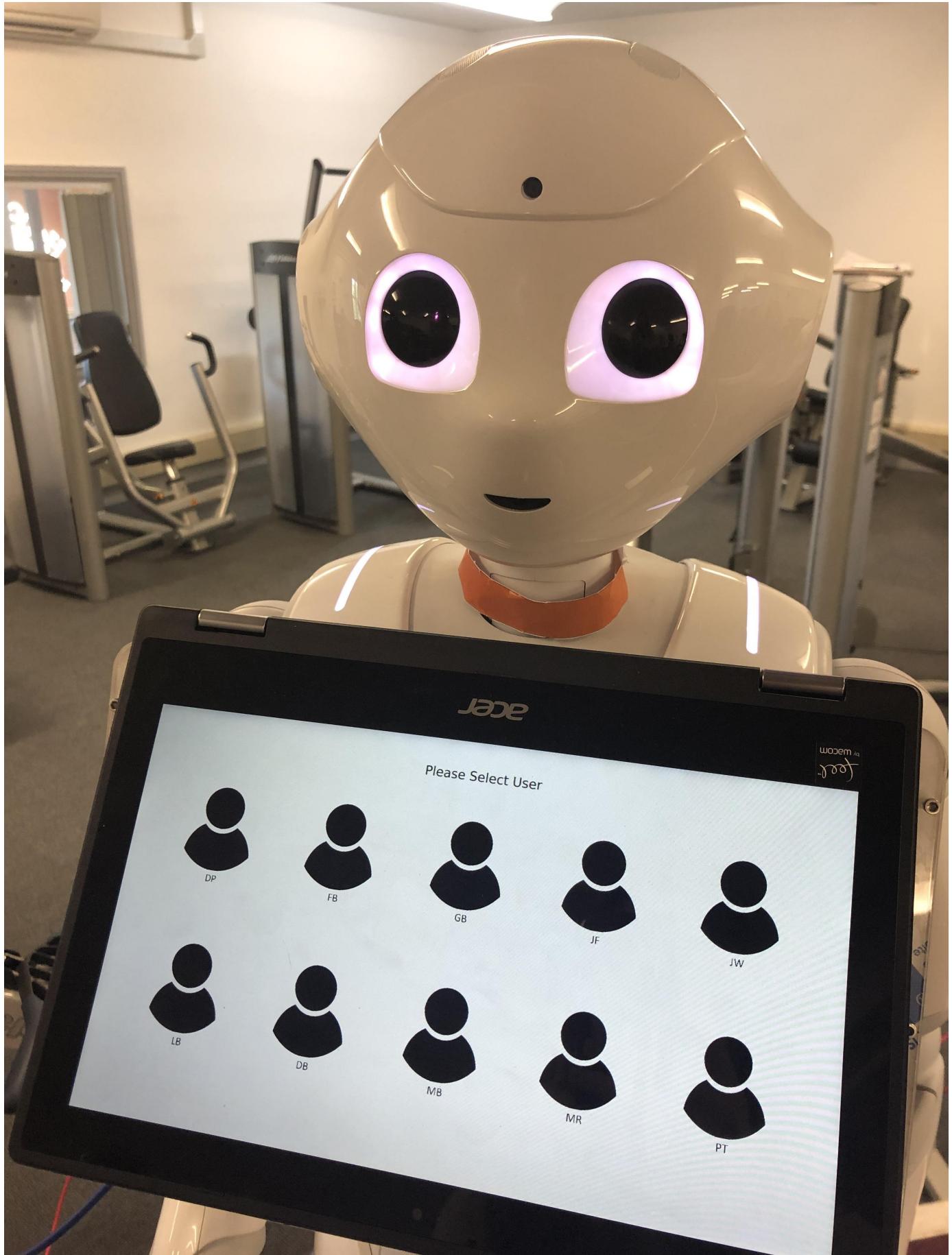
Please enter your programme User ID (your initials):

Robot A: Orange

Please think about your experiences with the Orange robot when answering these questions:

Robot A

Orange



As an exercise instructor, did you find the Orange robot to be:

Unexperienced



Experienced

Uninformed	Informed
Untrained	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Unqualified	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Unskilled	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Unintelligent	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Incompetent	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
Stupid	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

Did you feel that the Orange robot:

Didn't care about me	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Cared about me
Didn't have my interests at heart	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Had my interests at heart
Was self-centred	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Was not self-centred
Was not concerned with me	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Was concerned with me
Was insensitive	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Was sensitive
Was not understanding	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Was understanding

Did you find the Orange robot to be:

Irritable	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Good-natured
Gloomy	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Cheerful
Unfriendly	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Friendly

Please rate your impression of the Orange robot on the following scales:

Dislike	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Like
Unfriendly	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Friendly
Unkind	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Kind
Unpleasant	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Pleasant
Awful	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Nice

Please indicate your response to the following questions.

Over the course of this study, based on the time you've actually spent with the Orange robot:

Not at all Not much Not sure A bit A lot

	Not at all	Not much	Not sure	A bit	A lot
To what extent do you feel you developed a relationship with the Orange robot?	<input type="radio"/>				

To what extent do you feel the Orange robot developed a relationship with you?	<input type="radio"/>				
--	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Now hypothetically, if you were to work out with the Orange robot for a longer period, like you have been doing with the Purple robot:

	Not at all	Not much	Not sure	A bit	A lot
To what extent do you feel you could develop a relationship with the Orange robot?	<input type="radio"/>				
To what extent do you feel the Orange robot could develop a relationship with you?	<input type="radio"/>				

On seeing the Orange robot again for the first time in a while, do you feel its behaviour has changed since you saw it last? If so please give details on what differences you perceive and whether they might be positive/negative.

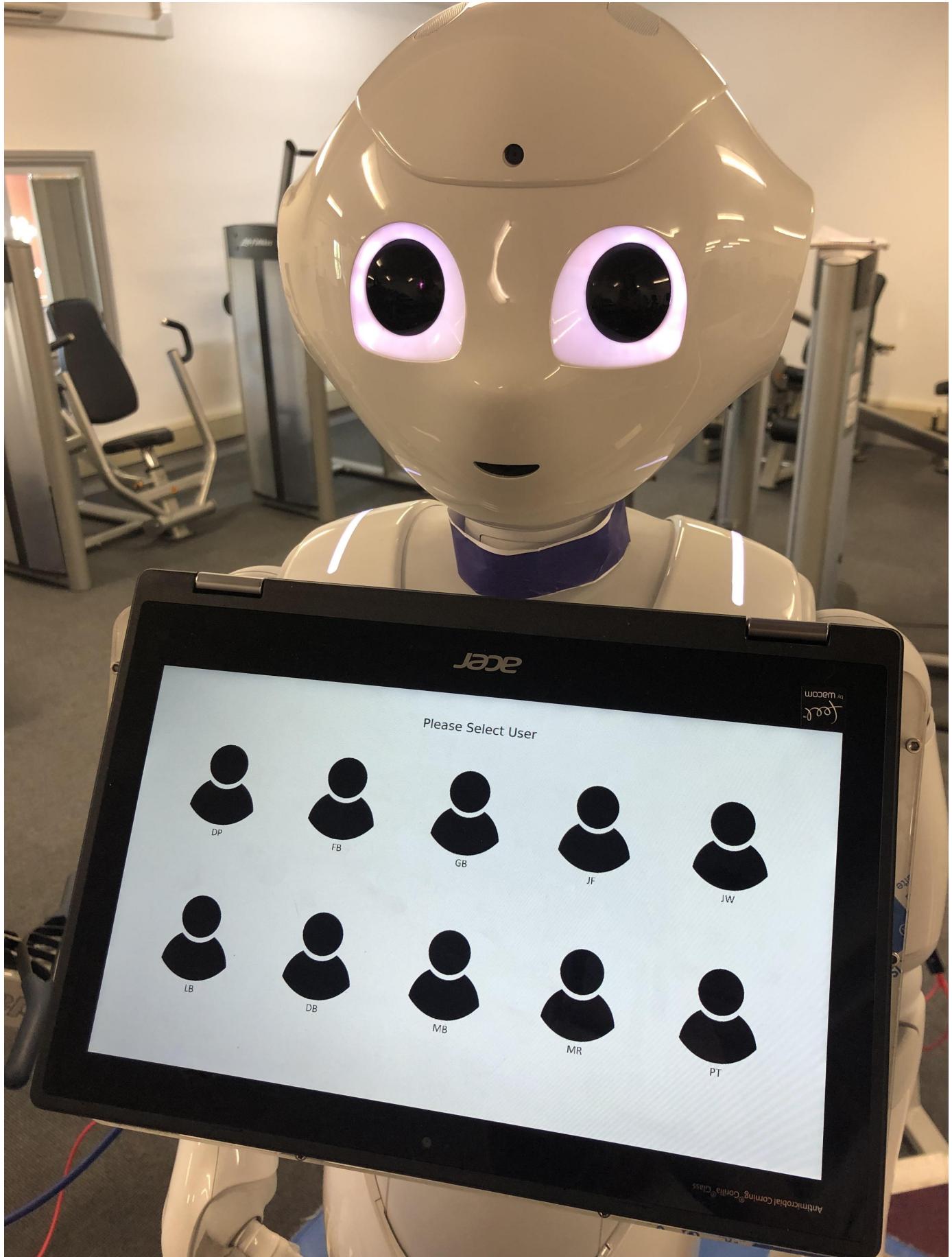
- Yes
- No
- Not Sure

Robot B: Purple

Please think about your experiences with the Purple robot when answering these questions:

Robot B

Purple



As an exercise instructor, did you find the Purple robot to be:

Unexperienced



Experienced

Uninformed	Informed
Untrained	<input type="radio"/>
Unqualified	<input type="radio"/>
Unskilled	<input type="radio"/>
Unintelligent	<input type="radio"/>
Incompetent	<input type="radio"/>
Stupid	<input type="radio"/>
	Trained
	Qualified
	Skilled
	Intelligent
	Competent
	Bright

Did you feel that the Purple robot:

Didn't care about me	<input type="radio"/>	Cared about me				
Didn't have my interests at heart	<input type="radio"/>	Had my interests at heart				
Was self-centred	<input type="radio"/>	Was not self-centred				
Was not concerned with me	<input type="radio"/>	Was concerned with me				
Was insensitive	<input type="radio"/>	Was sensitive				
Was not understanding	<input type="radio"/>	Was understanding				

Did you find the Purple robot to be:

Irritable	<input type="radio"/>	Good-natured				
Gloomy	<input type="radio"/>	Cheerful				
Unfriendly	<input type="radio"/>	Friendly				

Please rate your impression of the Purple robot on the following scales:

Dislike	<input type="radio"/>	Like				
Unfriendly	<input type="radio"/>	Friendly				
Unkind	<input type="radio"/>	Kind				
Unpleasant	<input type="radio"/>	Pleasant				
Awful	<input type="radio"/>	Nice				

Please indicate your response to the following questions:

	Not at all	Not much	Not sure	A bit	A lot
To what extent do you feel you developed a relationship with the Purple robot?	<input type="radio"/>				

Not at all Not much Not sure A bit A lot

To what extent do you feel the Purple robot developed a relationship with you?

Do you feel the Purple robot's behaviour has changed in any way over the course of the study, while you've been working out with it? If so please give details as to any changes you have perceived and whether they might be positive/negative.

- Yes
 No
 Not Sure

Robot A: Orange versus Robot B: Purple

Now thinking about comparing the two robots:

How would you describe the Orange and Purple robots as exercise instructors? What differences, if any, did you perceive between them?

Do you feel that one of the robots encouraged you to perform better than the other?

- Orange
 Purple
 No preference

Did you prefer working with one robot over the other?

- Orange
 Purple

No preference

If you were to undertake another similar exercise training programme, would you want to work with a robot like Pepper again? If so, would you have a preference for either the Orange or the Purple robot? Please leave any related reasoning in the boxes provided.

- I'd probably prefer not to work with a robot like this in future
- I'd like to work with the Purple robot again
- I'd like to work with the Orange robot again
- I'm not sure, or have no preference

Pre-Interview Thoughts

These questions are designed to capture your overall experience of taking part in this study. If you are willing and able to take part in a post-study interview then we may refer back to these answers during that interview for further discussion.

Please briefly summarise your experience of undertaking the Couch to 5km with Pepper and Don.

How would you describe the role of the robot with regards to the exercise programme? How does this compare to the role of the instructor (Don) and the researcher (Katie)?

How does your experience of this programme and working with Pepper compare to any other exercise you do or previous exercise programmes you may have completed? If you ever tried Couch to 5km with the smartphone app/podcast, how does our robot setup compare for you?

Finally, drawing on your experiences during this study, what are your thoughts on robot-supported exercise like this? For yourself during the programme, but also in the future or for others.

Please feel free (but not obliged) to leave any additional thoughts or comments here:

Powered by Qualtrics

08/07/2019 Monday #Week 4

(L) - Session 9 - Teacher - +

Walk = 2.8 ✓ 3 HR = 120 - 170 +
Run = 5.5 → 6 OK

* Lots of suggestions! # Probability

↳ Not bad though. # repeat.

↳ She knows what I want
to say!

Amazing Effort! #100% #Sweat.

Great Session. Form:

08/07

(JF) - Session 8 - Control - N ↫
Effortless → Good Running.

* Deeper Emotional assessment, 10

Sad? Tired? Unmotivated? PRE

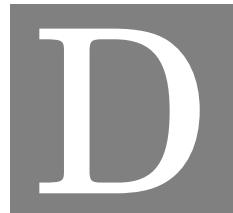
↳ Great Speed + Effort / ↑ HR ❤

50-50 / ^{way} too much Sympathy #Rewarded

OK



Form:



APPENDIX D

This appendix contains the following resources from the study with therapists as referred to in Chapter 5:

- Fisher's exact test results comparing the action distributions produced by each system, within-participant
- Fisher's exact test results comparing the action distributions produced by the first and second iterations of the Heuristic system, between-participant
- Fitness instructor's in-session notes from Phase 1 and Phase 3 testing of the Heuristic system
- Participant responses to the question regarding the role of the robot, fitness instructor and researcher
- Participant responses to the question regarding differences between the Heuristic and IML robots

APPENDIX D. APPENDIX D

		H1	IML-S2	H2	IML-A
LB	IML-S1	p <<0.001	p = 0.704	p <<0.001	p = 0.106
	H1	x	p <<0.001	p <<0.001	p <<0.001
	IMLS-2	x	x	p <<0.001	p = 0.803
	H2	x	x	x	p <<0.001
FB	IML-S1	p <<0.001	p = 0.654	p <<0.001	p = 0.001
	H1	x	p <<0.001	p <<0.001	p <<0.001
	IMLS-2	x	x	p <<0.001	p = 0.029
	H2	x	x	x	p <<0.001
DB	IML-S1	p <<0.001	p = 0.342	p <<0.001	p = 0.397
	H1	x	p <<0.001	p <<0.001	p <<0.001
	IMLS-2	x	x	p <<0.001	p <<0.001
	H2	x	x	x	p <<0.001
JF	IML-S1	p <<0.001	p = 0.816	p <<0.001	p = 0.016
	H1	x	p <<0.001	p <<0.001	p <<0.001
	IMLS-2	x	x	p <<0.001	p = 0.330
	H2	x	x	x	p <<0.001
MR	IML-S1	p <<0.001	p <<0.001	p <<0.001	p = 0.027
	H1	x	p <<0.001	p <<0.001	p <<0.001
	IMLS-2	x	x	p <<0.001	p = 0.581
	H2	x	x	x	p <<0.001
DP	IML-S1	p <<0.001	p = 0.015	p <<0.001	p <<0.001
	H1	x	p <<0.001	p <<0.001	p <<0.001
	IMLS-2	x	x	p <<0.001	p = 0.906
	H2	x	x	x	p <<0.001
JW	IML-S1	p <<0.001	p = 0.529	p <<0.001	p = 0.070
	H1	x	p <<0.001	p <<0.001	p <<0.001
	IMLS-2	x	x	p <<0.001	p = 0.776
	H2	x	x	x	p <<0.001
GB	IML-S1	p <<0.001	p = 0.331	p <<0.001	p <<0.001
	H1	x	p <<0.001	p <<0.001	p <<0.001
	IMLS-2	x	x	p <<0.001	p = 0.246
	H2	x	x	x	p <<0.001
PT	IML-S1	p <<0.001	p = 0.032	p <<0.001	p = 0.001
	H1	x	p <<0.001	p <<0.001	p <<0.001
	IMLS-2	x	x	p <<0.001	p = 0.107
	H2	x	x	x	p <<0.001

Table D.1: Fisher's exact test results comparing robot behaviour across conditions, within-participant according to the session data listed in Table 5.5. Non-significant results (representing the minority) are highlighted in bold.

H1 Fischer's Test	FB	DB	JF	MR	DP	JW	GB	PT
LB	p <<0.001	p <<0.001	p <<0.001	p = 0.363	p <<0.001	p = 0.473	p = 0.261	p = 0.002
FB	x		p <<0.001	p <<0.001	p = 0.007	p = 0.007	p = 0.003	p = 0.052
DB	x	x		p <<0.001	p <<0.001	p <<0.001	p <<0.001	p = 0.095
JF	x	x	x		p <<0.001	p <<0.001	p <<0.001	p <<0.001
MR	x	x	x		x	p <<0.001	p = 0.185	p = 0.052
DP	x	x	x		x		p <<0.001	p <<0.001
JW	x	x	x		x	x		p = 0.576
GB	x	x	x		x	x	x	p <<0.001

Table D.2: Fisher's exact test results comparing H1 robot behaviour across participants. *Non-significant* results (representing the minority) are highlighted in bold.

H2 Fischer's Test	FB	DB	JF	MR	DP	JW	GB	PT
LB	p = 0.006	p = 0.300	p = 0.856	p = 1.0	p <<0.001	p <<0.001	p <<0.001	p = 0.530
FB	x	p = 0.196	p - 0.004	p = 0.012	p <<0.001	p <<0.001	p <<0.001	p <<0.001
DB	x	x	p = 0.466	p = 0.279	p <<0.001	p <<0.001	p <<0.001	p = 0.018
JF	x	x	x	p = 0.755	p <<0.001	p <<0.001	p <<0.001	p = 0.130
MR	x	x	x	x	p <<0.001	p <<0.001	p <<0.001	p = 0.749
DP	x	x	x	x	x	p <<0.001	p = 1.0	p <<0.001
JW	x	x	x	x	x	x	p <<0.001	p <<0.001
GB	x	x	x	x	x	x	x	p <<0.001

Table D.3: Fisher's exact test results comparing H2 robot behaviour across participants. *Non-significant* results (representing the minority) are highlighted in bold.

APPENDIX D. APPENDIX D

Session	Fitness Instructor Notes
LB 3	Warm up still a little quiet.
LB 5	Very good suggestions. Pepper's suggestions might not be what *I* would say in that exact same situation, however it doesn't mean that what was said or suggested was 'wrong'
FB 1	Suggesting challenge/praise actions a little too often which could be repetitive, however as a whole made great decisions. Impressed.
FB 3	Lots of praise. Not too much interaction between FB and Pepper. Some repetitive speech. As a whole, great session FB worked hard and Pepper made some good action choices.
DB 2	Good warm up behaviour. Slightly repetitive. No sass.
DB 4	Fairly good suggestions but could be more challenging.
JF 2	Hmm not so good. Very repetitve speech. Hard-coded responses. More variation needed.
JF 4	Over-using sympathise action. Too much repetition (don't think client realised).
MR 1	Great session.
MR 3	Fairly good speech choice, sympathetic and praise.
DP 2	Pepper did not suggest action styles although actions spoken were very good.
DP 4	Far too much maintenance speech. Conversation not flowing so well.
JW 2	Not too much interaction w/ Pepper. Didn't need too much speech but occasionally repetitive speech.
JW 4	A 'standard' training session - averagely good, no complaints. Pepper made good varied speech. JW could have been challenged more.
GB 1	Very good [H] suggestions - same action suggestion at the same time as me! Sometimes repetitive speech. As a whole, good session.
GB 3	Pepper = ok suggestions. Maintenance = a little repetitive.
PT 1	Quiet on warm up, not much variation.
PT 3	Average/ok speech choices. Could be more engaging.

Table D.4: Fitness instructor notes taken during a subset of heuristic sessions from Phase 1 testing as per Table 5.5.

Session	Fitness Instructor Notes
LB 25	Action = ok. Speech/action choice has variation but a random mix of challenge and sympathy... contradictory. Mainly challenging = good. Push him. A *lot* of speech, no opportunity to zone out. Overall the [H] is doing alright/not bad. Just very talkative that could result in repeated speech. Although happy with the check pre. Cool down sympathy was good but needs work as it kinda breaks the immersion.
FB 24	Actions = starting actions = pretty good! Getting the client in a challenge oriented mind-set. 5 min. in tells client to zone-out and take the mind where it needs to go to keep that intensity. After that, I would have done less-frequent speech i.e. back off to let the client follow those orders and zone out without distraction. [H] Pepper is very challenging but FB can handle it. Good action/dialogue variation and check pre. Repetitive speech starting to show (10 mins left). No speech evolution, no real/noticeable change in behaviour over the run i.e. end of run push.
DB 21	Actions = very sympathetic. A lot of constant speech. DB not happy with speech! Let her zone out! No escape from the repetition. Talking not helping, This is not a good session, opposite effect, de-motivating. Walking, Pepper too off-putting to focus. Worst session ever.
JF 25	Actions = very challenging from the start. Each [H] acts quite differently for each client. Showing variation but not as desirable... not so good choice or timing of certain actions/speech. [H] goes through a lot of speech, something every 30 seconds for 30 minutes can be a bit much, given the limited dialogue. Displaying a noticeable amount of repeated speech. Not bad (ok). Cool down speech off.
MR 21	Actions = very challenging Pepper from the start - believe me Pepper, she's giving you serious energy and intensity! Good choice and variation of actions. Very challenging, better than being too over-positive but not really acknowledging the client's effort - praise! Just still more repeated speech, no evolution. End of run = so-so, nothing special.
DP 27	Action = 1st action = 'slow down' seconds in, and 'don't forget to pace yourself' - no reason for this! Luckily DP not listening to Pepper and still increasing speed (although more slowly/w. less confidence). Speed down only, repeated. Not impressed for DP's last session. DP not taking the easy option, and still talking to Pepper. I'm not happy with [H].
JW 25	Actions = good start! Relevant choice and variation. Hard to see/predict [H] behaviour. Very talkative! Not providing much time/opportunity for the client to zone out. Talking so often in repetition it may be easy to zone-out Pepper i.e. ignoring her = unvalued speech. A *lot* of repeated speech. Lacking connection. No evolution or change in behavioural actions over the course of the run. Cool down speech off.
GB 26	Only speed down! *NOT* quite what I'm looking for in a trainer...
PT 21	Action = good timing, very challenging. Not identifying client struggle, kept them pushing through.

Table D.5: Fitness instructor notes taken during each participant's session with the updated heuristic robot in Phase 3 testing as per Table 5.5.

APPENDIX D. APPENDIX D

	How would you describe the role of the robot with regards to the exercise programme? How does this compare to the role of the instructor (Don) and the researcher (Katie)?
LB	Pepper was a good instructor and positively motivated my runs. The role of Don assisted this in that having him there meant I could follow the robot's instructions safe in the knowledge that there was some support there should anything go wrong! Katie was informative about project throughout.
FB	The robot delivered the exercise programme, having been pre-programmed. Don was there to observe and Katie organised the study. Don and Katie's roles were more fluid - giving advice/answering questions and rescheduling sessions as necessary, whereas the robot felt quite fixed in what it could do and say as it led each run.
DB	I felt positively obliged to show up because I want Katie to have a good, meaningful piece of work. I wanted to better myself and get back on track and I felt such a regime could help. I really enjoyed speaking with Don but also I would defo sign up to PT sessions with him if that were a thing (rather than Pepper, as Pepper is not for me).
JF	Pepper was really helpful in the beginning to tell me when to run and walk- later Pepper was useful in helping me keep focus (breathing, shoulders) and to break up the running with motivation. Don's presence overall was also really motivating as he was really encouraging after each run and with the overall programme. Katie wasn't at the sessions very often so more of a background presence in the programme.
MR	Pepper had the role of keeping time. Sometimes Pepper motivated me to give more energy at the end of a running phase
DP	I did not see much of Katie after the first weeks. Pepper was important in getting me to run and walk and slowly build up to running for 30 minutes. Don's stretching routines were essential in my commitment as a major fear was dropping out because of a pulled muscle.
JW	The robot was vital mainly because it timed my runs and gave some occasional useful comments. However, Don and Katie were also vital in providing encouragement and incentive after each session. I think if it had just been Pepper and me and no other human involved I may not have completed the programme...i.e. Pepper switches off when the treadmill stops but my concerns/aches/pains don't!
GB	Even being someone that likes working out how things work (and being a roboticist as well) it was hard to say exactly what was going on behind the scenes. But, on the face of it, it would appear that Pepper was deciding what comments to make at which points, and Don was supervising or monitoring it in some way. Whether Don had any sort of direct control over the robot's comments remains unclear to me.
PT	The robot in my opinion did a great job in helping me achieve the couch to 5k programme. It was a great gym-buddy companion that maked me wanting to go to the session and try my best. It is quite hard to explain how I felt in writing as for examples sometimes in the gym or in a personal training session you don't feel like talking and you just want to exercise in order to free your mind a bit, during those days Pepper was ideal! However there are some other days that you need the extra push from your personal trainer because you just want to give up, at those times I felt that Pepper could have done more to encourage me or to give me actual feedback that will push me to work harder. I can imagine a gym with Pepper-like instructors next to treadmills but of course with the presence of experienced human instructors there to take care of you in case that something goes wrong.

Table D.6: Participant descriptions of robot, fitness instructor and researcher role in the context of delivering the couch to 5km programme.

	How would you describe the [H] and [IML] robots as exercise instructors? What differences, if any, did you perceive between them?
LB	[H] seemed more talkative although that may just be based on today's session seeing her again for the first time in a while.
FB	They were both ok as exercise instructors. I prefer the interaction you get with human exercise instructors. Both of the robots' phrases can get quite repetitive, however I felt I pushed myself harder (particularly with the [IML] robot) than I would have done on my own. They definitely provided motivation to keep going and had some knowledge too on how technically to run (I heard more phrases from the [IML] robot due to working with her more). They seemed friendly, particularly when saying hello and goodbye, which made it easier than expected to start forming a relationship with them (referring to Pepper as her etc.)
DB	[H] is super annoying and I feel for all participants who interacted with that thing. [IML] was nice and quiet and let me get on with it and tbh towards the end I ignored what it was saying- could not even tell you if the text was the same for the past two weeks... but when you compare it to orange- goodness gracious [IML] is amazing and seems an amazing companion (where I am ambivalent towards my feelings for Pepper)
JF	I think the [H] robot focussed more on trying to get people to put in the most or more effort whereas the [IML] robot was more gentle and wanted people to just try the best they could
MR	I feel like the [H] Pepper is more wholesome and more concerned with my well being. The [IML] Pepper pushed me harder, which I enjoyed
DP	[IML] had more to say and with less repetition per session. [H] seemed overly cautious. I preferred [IML].
JW	I always found it difficult to distinguish any differences. However, [H] robot today seemed to give more comments that were based around by wellbeing whereas [IML] would give more comments on working harder. I also found that [IML] would often make a comment that wasn't fit for that part of the session...i.e. telling me to push myself when walking or telling me I'm nearly finished when I've only just started!
GB	I would describe [IML] as motivational/instructive as it usually commented with encouragement or occasionally running tips. From the limited experience I got of [H], I would describe it as ultra-careful, repetitive and a bit tedious.
PT	[IML] robot made the experience more personal as it was using my name when it was saying encouraging stuff. [H] robot felt more like a finely tuned piece of equipment, which was giving instructions on what to do and saying stuff automatically without taking into account what I was actually doing.

Table D.7: Participant descriptions of the IML and H robots and any differences between them.

BIBLIOGRAPHY

- Abbeel, P. & Ng, A. Y. (2004), Apprenticeship learning via inverse reinforcement learning, in ‘Proceedings of the Twenty-First International Conference on Machine Learning’, ICML ’04, Association for Computing Machinery, Banff, Alberta, Canada, p. 1.
- Aristotle (1954), *Rhetoric*, w. r. roberts, trans. edn, Random House, New York.
- Arkin, R. C., Fujita, M., Takagi, T. & Hasegawa, R. (2001), Ethological modeling and architecture for an entertainment robot, in ‘Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)’, Vol. 1, pp. 453–458 vol.1.
- Armitage, C. J. & Conner, M. (2001), ‘Efficacy of the Theory of Planned Behaviour: A meta-analytic review’, *British Journal of Social Psychology* **40**(4), 471–499.
- Aronson, E., Wilson, T. D. & Akert, R. M. (2010), *Social Psychology*, Prentice Hall, Upper Saddle River, NJ.
- Azenkot, S., Feng, C. & Cakmak, M. (2016), Enabling building service robots to guide blind people a participatory design approach, in ‘2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)’, pp. 3–10.
- Baltrušaitis, T., Zadeh, A., Lim, Y. C. & Morency, L.-P. (2018), OpenFace 2.0: Facial Behavior Analysis Toolkit, in ‘2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)’, pp. 59–66.
- Bartneck, C., Kulić, D., Croft, E. & Zoghbi, S. (2009), ‘Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots’, *International journal of social robotics* **1**(1), 71–81.
- Beer, J. M., Smarr, C.-A., Chen, T. L., Prakash, A., Mitzner, T. L., Kemp, C. C. & Rogers, W. A. (2012), The Domesticated Robot: Design Guidelines for Assisting Older Adults to Age in Place, in ‘Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction’, HRI ’12, ACM, New York, NY, USA, pp. 335–342.
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E. & Fowler, J. H. (2012), ‘A 61-million-person experiment in social influence and political mobilization’, *Nature* **489**(7415), 295–298.

BIBLIOGRAPHY

- Bonner, N. S., O'Halloran, P. D., Bernhardt, J. & Cumming, T. B. (2016), 'Developing the Stroke Exercise Preference Inventory (SEPI)', *PLOS ONE* **11**(10), e0164120.
- Brehm, J. W. (1966), *A Theory of Psychological Reactance*, A Theory of Psychological Reactance, Academic Press, Oxford, England.
- BSI (2016), 'BS 8611:2016 Robots and robotic devices: Guide to the ethical design and application of robots and robotic systems'.
- Cacioppo, J. T. & Petty, R. E. (1984), 'The Elaboration Likelihood Model of Persuasion', *ACR North American Advances* **NA-11**.
- Cañamero, L. & Lewis, M. (2016), 'Making New "New AI" Friends: Designing a Social Robot for Diabetic Children from an Embodied AI Perspective', *International Journal of Social Robotics* **8**(4), 523–537.
- Chan, J. & Nejat, G. (2012), 'Social Intelligence for a Robot Engaging People in Cognitive Training Activities', *International Journal of Advanced Robotic Systems* **9**(4), 113.
- Chang, W.-L. & Šabanović, S. (2015), Interaction Expands Function: Social Shaping of the Therapeutic Robot PARO in a Nursing Home, in 'Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction', HRI '15, ACM, New York, NY, USA, pp. 343–350.
- Chidambaram, V., Chiang, Y.-H. & Mutlu, B. (2012), Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues, in 'Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction', ACM, pp. 293–300.
- Clark-Turner, M. & Begum, M. (2018), Deep Reinforcement Learning of Abstract Reasoning from Demonstrations, in 'Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction', HRI '18, Association for Computing Machinery, Chicago, IL, USA, pp. 160–168.
- Cruz-Maya, A. & Tapus, A. (2018), Negotiating with a Robot: Analysis of Regulatory Focus Behavior - IEEE Conference Publication, in '2018 IEEE International Conference on Robotics and Automation (ICRA)'.
- Curry, L. (2015), 'Fundamentals of Qualitative Research Methods'.
- de Graaf, M. M. A., Allouch, S. B. & Klamer, T. (2015), 'Sharing a life with Harvey: Exploring the acceptance of and relationship-building with a social robot', *Computers in Human Behavior* **43**, 1–14.

- de Vries, R. A. J., Truong, K. P., Zaga, C., Li, J. & Evers, V. (2017), 'A word of advice: How to tailor motivational text messages based on behavior change theory to personality and gender', *Personal and Ubiquitous Computing* **21**(4), 675–687.
- Department of Health (2010), 'Healthy Foundations Life-stage Segmentation Model Toolkit'.
- Desroches, S., Lapointe, A., Ratté, S., Gravel, K., Légaré, F. & Turcotte, S. (2013), Interventions to enhance adherence to dietary advice for preventing and managing chronic diseases in adults, in 'Cochrane Database of Systematic Reviews', John Wiley & Sons, Ltd.
- Feil-Seifer, D. & Matarić, M. J. (2005), Defining socially assistive robotics, in 'Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference On', IEEE, pp. 465–468.
- Fitter, N. T., Mohan, M., Kuchenbecker, K. J. & Johnson, M. J. (2020), 'Exercising with Baxter: Preliminary support for assistive social-physical human-robot interaction', *Journal of NeuroEngineering and Rehabilitation* **17**(1), 19.
- Fong, T., Nourbakhsh, I. & Dautenhahn, K. (2003), 'A survey of socially interactive robots', *Robotics and Autonomous Systems* **42**(3–4), 143–166.
- Forkan, R., Pumper, B., Smyth, N., Wirkkala, H., Cirol, M. A. & Shumway-Cook, A. (2006), 'Exercise adherence following physical therapy intervention in older adults with impaired balance', *Physical Therapy* **86**(3), 401–410.
- Forlizzi, J., DiSalvo, C. & Gemperle, F. (2004), 'Assistive Robotics and an Ecology of Elders Living Independently in Their Homes', *Hum.-Comput. Interact.* **19**(1), 25–59.
- Gale, N. K., Heath, G., Cameron, E., Rashid, S. & Redwood, S. (2013), 'Using the framework method for the analysis of qualitative data in multi-disciplinary health research', *BMC Medical Research Methodology* **13**, 117.
- Gass, R. H. (2015), Social Influence, Sociology of, in J. D. Wright, ed., 'International Encyclopedia of the Social & Behavioral Sciences (Second Edition)', Elsevier, Oxford, pp. 348–354.
- Gass, R. H. & Seiter, J. S. (2015), *Persuasion: Social Influence and Compliance Gaining*, Routledge.
- Ghazali, A. S., Ham, J., Barakova, E. & Markopoulos, P. (2019), 'Assessing the effect of persuasive robots interactive social cues on users' psychological reactance, liking, trusting beliefs and compliance: Advanced Robotics: Vol 33, No 7-8', *Advanced Robotics* **33**, 325–337.
- Gillet, S. & Leite, I. (2020), A Robot Mediated Music Mixing Activity for Promoting Collaboration among Children, in 'Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction', HRI '20, Association for Computing Machinery, Cambridge, United Kingdom, pp. 212–214.

BIBLIOGRAPHY

- Gockley, R. & Mataric, M. J. (2006), Encouraging physical therapy compliance with a hands-off mobile robot, *in* ‘Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction’, ACM, pp. 150–155.
- González, J. C., Pulido, J. C., Fernández, F. & Suárez-Mejías, C. (2015), Planning, execution and monitoring of physical rehabilitation therapies with a robotic architecture., *in* ‘MIE’, pp. 339–343.
- Gosling, S. D., Rentfrow, P. J. & Swann, W. B. (2003), ‘A very brief measure of the Big-Five personality domains’, *Journal of Research in Personality* **37**(6), 504–528.
- Greczek, J., Short, E., Clabaugh, C. E., Swift-Spong, K. & Mataric, M. (2014), Socially Assistive Robotics for Personalized Education for Children, *in* ‘AAAI Fall Symposium on Artificial Intelligence and Human-Robot Interaction (AI-HRI)’.
- Green, A., Huttenrauch, H., Norman, M., Oestreicher, L. & Eklundh, K. S. (2000), User centered design for intelligent service robots, *in* ‘Proceedings 9th IEEE International Workshop on Robot and Human Interactive Communication. IEEE RO-MAN 2000 (Cat. No.00TH8499)’, pp. 161–166.
- Hall, J., Tritton, T., Rowe, A., Pipe, A., Melhuish, C. & Leonards, U. (2014), ‘Perception of own and robot engagement in human–robot interactions and their dependence on robotics knowledge’, *Robotics and Autonomous Systems* **62**(3), 392–399.
- Ham, J., Cuijpers, R. H. & Cabibihan, J.-J. (2015), ‘Combining Robotic Persuasive Strategies: The Persuasive Power of a Storytelling Robot that Uses Gazing and Gestures’, *International Journal of Social Robotics* **7**(4), 479–487.
- Ham, J. & Midden, C. J. H. (2014), ‘A Persuasive Robot to Stimulate Energy Conservation: The Influence of Positive and Negative Social Feedback and Task Similarity on Energy-Consumption Behavior’, *International Journal of Social Robotics* **6**(2), 163–171.
- Jenkins, S. & Draper, H. (2015), ‘Care, Monitoring, and Companionship: Views on Care Robots from Older People and Their Carers’, *International Journal of Social Robotics* **7**(5), 673–683.
- Jones, S. (1992), ‘Was There a Hawthorne Effect?’, *American Journal of Sociology* **98**(3), 451–468.
- Jordan, J. L., Holden, M. A., Mason, E. E. & Foster, N. E. (2010), Interventions to improve adherence to exercise for chronic musculoskeletal pain in adults, *in* ‘Cochrane Database of Systematic Reviews’, John Wiley & Sons, Ltd.
- Kahn, Jr., P. H., Kanda, T., Ishiguro, H., Gill, B. T., Shen, S., Gary, H. E. & Ruckert, J. H. (2015), Will People Keep the Secret of a Humanoid Robot?: Psychological Intimacy in HRI,

- in 'Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction', HRI '15, ACM, New York, NY, USA, pp. 173–180.
- Kang, K. I., Freedman, S., Mataric, M. J., Cunningham, M. J. & Lopez, B. (2005), A hands-off physical therapy assistance robot for cardiac patients, *in* '9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.', pp. 337–340.
- Karmali, K. N., Davies, P., Taylor, F., Beswick, A., Martin, N. & Ebrahim, S. (2014), Promoting patient uptake and adherence in cardiac rehabilitation, *in* 'Cochrane Database of Systematic Reviews', John Wiley & Sons, Ltd.
- Kelman, H. C. (1958), 'Compliance, Identification, and Internalization: Three Processes of Attitude Change', *The Journal of Conflict Resolution* **2**(1), 51–60.
- Knox, W. B., Spaulding, S. & Breazeal, C. (2014), Learning Social Interaction from the Wizard: A Proposal, *in* 'Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence'.
- Lara, J. S., Casas, J., Aguirre, A., Munera, M., Rincon-Roncancio, M., Irfan, B., Senft, E., Belpaeme, T. & Cifuentes, C. A. (2017), Human-robot sensor interface for cardiac rehabilitation, *in* '2017 International Conference on Rehabilitation Robotics (ICORR)', pp. 1013–1018.
- Lee, H. R., Sabanovic, S., Chang, W.-L., Nagata, S., Piatt, J., Bennett, C. & Hakken, D. (2017), Steps Toward Participatory Design of Social Robots: Mutual Learning with Older Adults with Depression, *in* 'Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction', HRI '17, ACM, New York, NY, USA, pp. 244–253.
- Lee, S. A. & Liang, Y. J. (2019), 'Robotic foot-in-the-door: Using sequential-request persuasive strategies in human-robot interaction - ScienceDirect', *Computers in Human Behavior* **90**, 351–356.
- Leite, I., Pereira, A., Castellano, G., Mascarenhas, S., Martinho, C. & Paiva, A. (2011), Modelling Empathy in Social Robotic Companions, *in* 'Advances in User Modeling', Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp. 135–147.
- Lemaignan, S., Jacq, A., Hood, D., Garcia, F., Paiva, A. & Dillenbourg, P. (2016), 'Learning by Teaching a Robot: The Case of Handwriting', *IEEE Robotics Automation Magazine* **23**(2), 56–66.
- Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A. & Alami, R. (2017), 'Artificial cognition for social human–robot interaction: An implementation', *Artificial Intelligence* **247**, 45–69.
- Leme, B., Hirokawa, M., Kadone, H. & Suzuki, K. (2019), 'A Socially Assistive Mobile Platform for Weight-Support in Gait Training', *International Journal of Social Robotics* .

BIBLIOGRAPHY

- Lohani, M., Stokes, C., McCoy, M., Bailey, C. A. & Rivers, S. E. (2016), Social Interaction Moderates Human-Robot Trust-Reliance Relationship and Improves Stress Coping, in 'The Eleventh ACM/IEEE International Conference on Human Robot Interaction', HRI '16, IEEE Press, Piscataway, NJ, USA, pp. 471–472.
- Louie, W. Y. G., Li, J., Vaquero, T. & Nejat, G. (2014), A focus group study on the design considerations and impressions of a socially assistive robot for long-term care, in 'The 23rd IEEE International Symposium on Robot and Human Interactive Communication', pp. 237–242.
- Louie, W.-Y. G. & Nejat, G. (2020), 'A Social Robot Learning to Facilitate an Assistive Group-Based Activity from Non-expert Caregivers', *International Journal of Social Robotics*.
- Lucas, G. M., Boberg, J., Traum, D., Artstein, R., Gratch, J., Gainer, A., Johnson, E., Leuski, A. & Nakano, M. (2018), Getting to Know Each Other: The Role of Social Dialogue in Recovery from Errors in Social Robots, in 'Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction', HRI '18, ACM, New York, NY, USA, pp. 344–351.
- Malik, N. A., Hanapiyah, F. A., Rahman, R. A. A. & Yussof, H. (2016), 'Emergence of Socially Assistive Robotics in Rehabilitation for Children with Cerebral Palsy: A Review', *International Journal of Advanced Robotic Systems* **13**(3), 135.
- Marilyn B. Cole MS, OTR/L, F. & Valnere McLean MS, BCN, O. F. (2003), 'Therapeutic Relationships Re-Defined', *Occupational Therapy in Mental Health* **19**(2), 33–56.
- Martelaro, N., Nneji, V. C., Ju, W. & Hinds, P. (2016), Tell Me More: Designing HRI to Encourage More Trust, Disclosure, and Companionship, in 'The Eleventh ACM/IEEE International Conference on Human Robot Interaction', HRI '16, IEEE Press, Piscataway, NJ, USA, pp. 181–188.
- Martinez-Martin, E. & Cazorla, M. (2019), 'A Socially Assistive Robot for Elderly Exercise Promotion', *IEEE Access* **7**, 75515–75529.
- McCroskey, J. C. & Teven, J. J. (1999), 'Goodwill: A reexamination of the construct and its measurement', *Communication Monographs* **66**(1), 90–103.
- McCroskey, J. C. & Young, T. J. (1981), 'Ethos and credibility: The construct and its measurement after three decades', *Central States Speech Journal* **32**(1), 24–34.
- Mead, R., Wade, E., Johnson, P., Clair, A. S., Chen, S. & Mataric, M. J. (2010), An architecture for rehabilitation task practice in socially assistive human-robot interaction, in '19th International Symposium in Robot and Human Interactive Communication', pp. 404–409.
- Milgram, S. (1974), *Obedience to Authority: An Experimental View*, Harper & Row.

- Nakagawa, K., Shiomi, M., Shinozawa, K., Matsumura, R., Ishiguro, H. & Hagita, N. (2011), Effect of robot's active touch on people's motivation, *in* 'Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference On', IEEE, pp. 465–472.
- O'Keefe, D. J. (2002), *Persuasion: Theory and Research*, SAGE.
- O'Shea, S. D., Taylor, N. F. & Paratz, J. D. (2007), ' . . . But watch out for the weather: Factors affecting adherence to progressive resistance exercise for persons with COPD', *Journal of Cardiopulmonary Rehabilitation and Prevention* **27**(3), 166–174; quiz 175–176.
- Petty, R. E. & Cacioppo, J. T. (1984), 'Source Factors and the Elaboration Likelihood Model of Persuasion', *ACR North American Advances* **NA-11**.
- Pollock, A., Farmer, S. E., Brady, M. C., Langhorne, P., Mead, G. E., Mehrholz, J. & van Wijck, F. (2014), Interventions for improving upper limb function after stroke, *in* 'Cochrane Database of Systematic Reviews', John Wiley & Sons, Ltd.
- Pollock, A., Gray, C., Culham, E., Durward, B. R. & Langhorne, P. (2014), Interventions for improving sit-to-stand ability following stroke, *in* 'Cochrane Database of Systematic Reviews', John Wiley & Sons, Ltd.
- Pornpitakpan, C. (2004), 'The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence', *Journal of Applied Social Psychology* **34**(2), 243–281.
- Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R. & Ng, A. Y. (2009), Ros: an open-source robot operating system, *in* 'ICRA workshop on open source software', Vol. 3, Kobe, Japan, p. 5.
- Rescorla, R. A. (1971), 'Variation in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning', *Learning and Motivation* **2**(2), 113–123.
- Riek, L. D. (2012), 'Wizard of Oz studies in HRI: A systematic review and new reporting guidelines', *Journal of Human-Robot Interaction* **1**(1), 119–136.
- Rossi, S. & D'Alterio, P. (2017), Gaze Behavioral Adaptation Towards Group Members for Providing Effective Recommendations | SpringerLink, *in* 'International Conference on Social Robotics'.
- Sabanovic, S. (2010), 'Robots in Society, Society in Robots', *International Journal of Social Robotics* **2**(4), 439–450.
- Sabelli, A. M., Kanda, T. & Hagita, N. (2011), A conversational robot in an elderly care center: An ethnographic study, *in* '2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)', pp. 37–44.

BIBLIOGRAPHY

- Salam, H., Celiktutan, O., Hupont, I., Gunes, H. & Chetouani, M. (2017), 'Fully Automatic Analysis of Engagement and Its Relationship to Personality in Human-Robot Interactions', *IEEE Access* **5**, 705–721.
- Salem, M., Lakatos, G., Amirabdollahian, F. & Dautenhahn, K. (2015), Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust, in 'Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction', HRI '15, ACM, New York, NY, USA, pp. 141–148.
- Saunderson, S. & Nejat, G. (2019), 'It Would Make Me Happy if You Used My Guess: Comparing Robot Persuasive Strategies in Social Human–Robot Interaction - IEEE Journals & Magazine', *IEEE Robotics and Automation Letters* **4**, 1707–1714.
- Schneider, S., Goerlich, M. & Kummert, F. (2017), 'A framework for designing socially assistive robot interactions', *Cognitive Systems Research* **43**(Supplement C), 301–312.
- Schneider, S. & Kummert, F. (2020), 'Comparing Robot and Human guided Personalization: Adaptive Exercise Robots are Perceived as more Competent and Trustworthy', *INTERNATIONAL JOURNAL OF SOCIAL ROBOTICS* .
- Senft, E., Lemaignan, S., Baxter, P. E., Bartlett, M. & Belpaeme, T. (2019), 'Teaching robots social autonomy from in situ human guidance', *Science Robotics* **4**(35).
- Sequeira, P., Alves-Oliveira, P., Ribeiro, T., Di Tullio, E., Petisca, S., Melo, F. S., Castellano, G. & Paiva, A. (2016), Discovering social interaction strategies for robots from restricted-perception Wizard-of-Oz studies, in '2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)', pp. 197–204.
- Shao, M., Alves, S. F. D. R., Ismail, O., Zhang, X., Nejat, G. & Benhabib, B. (2019), You Are Doing Great! Only One Rep Left: An Affect-Aware Social Robot for Exercising, in '2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)', pp. 3811–3817.
- Sherif, M. & Hovland, C. I. (1961), *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*, Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change, Yale Univer. Press, Oxford, England.
- Shiomi, M., Kanda, T., Howley, I., Hayashi, K. & Hagita, N. (2015), 'Can a Social Robot Stimulate Science Curiosity in Classrooms?', *International Journal of Social Robotics* **7**, 1–12.
- Siegel, M., Breazeal, C. & Norton, M. I. (2009), Persuasive Robotics: The influence of robot gender on human behavior, in '2009 IEEE/RSJ International Conference on Intelligent Robots and Systems', pp. 2563–2568.

- Simons, H. W., Berkowitz, N. N. & Moyer, R. J. (1970), ‘Similarity, credibility, and attitude change: A review and a theory’, *Psychological Bulletin* **73**(1), 1–16.
- Singh, A., Klapper, A., Jia, J., Fidalgo, A., Tajadura-Jiménez, A., Kanakam, N., Bianchi-Berthouze, N. & Williams, A. (2014), Motivating People with Chronic Pain to Do Physical Activity: Opportunities for Technology Design, in ‘Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems’, CHI ’14, ACM, New York, NY, USA, pp. 2803–2812.
- Solman, B. & Clouston, T. (2016), ‘Occupational therapy and the therapeutic use of self’, *British Journal of Occupational Therapy* **79**(8), 514–516.
- Stanton, C. & Stevens, C. J. (2014), Robot Pressure: The Impact of Robot Eye Gaze and Lifelike Bodily Movements upon Decision-Making and Trust, in ‘Social Robotics’, Lecture Notes in Computer Science, Springer, Cham, pp. 330–339.
- Sussenbach, L., Riether, N., Schneider, S., Berger, I., Kummert, F., Lutkebohle, I. & Pitsch, K. (2014), A robot as fitness companion: Towards an interactive action-based motivation model, in ‘The 23rd IEEE International Symposium on Robot and Human Interactive Communication’, pp. 286–293.
- Swift-Spong, K., Short, E., Wade, E. & Mataric, M. J. (2015), ‘Effects of comparative feedback from a Socially Assistive Robot on self-efficacy in post-stroke rehabilitation’.
- Tapus, A. & Mataric, M. J. (2008), Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance., in ‘AAAI Spring Symposium: Emotion, Personality, and Social Behavior’, pp. 133–140.
- Tapus, A., Tapus, C. & Mataric, M. (2009), The role of physical embodiment of a therapist robot for individuals with cognitive impairments, in ‘Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium On’, IEEE, pp. 103–107.
- Tapus, A., Tăpuş, C. & Matarić, M. J. (2008), ‘User-robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy’, *Intelligent Service Robotics* **1**(2), 169–183.
- Taylor, R. R., Lee, S. W., Kielhofner, G. & Ketkar, M. (2009), ‘Therapeutic Use of Self: A Nationwide Survey of Practitioners’ Attitudes and Experiences’, *American Journal of Occupational Therapy* **63**(2), 198–207.
- Vandemeulebroucke, T., Dierckx de Casterlé, B. & Gastmans, C. (2018), ‘The use of care robots in aged care: A systematic review of argument-based ethics literature’, *Archives of Gerontology and Geriatrics* **74**, 15–25.

BIBLIOGRAPHY

- Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. (2003), 'User Acceptance of Information Technology: Toward a Unified View', *MIS Quarterly* **27**(3), 425–478.
- Visser, M., Brychta, R. J., Chen, K. Y. & Koster, A. (2014), 'Self-Reported Adherence to the Physical Activity Recommendation and Determinants of Misperception in Older Adults', *Journal of Aging and Physical Activity* **22**(2), 226–234.
- Wainer, J., Feil-Seifer, D. J., Shell, D., Mataric, M. J. et al. (2007), Embodiment and human-robot interaction: A task-based perspective, in 'Robot and Human Interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium On', IEEE, pp. 872–877.
- Wankel, L. M. (1985), 'Personal and Situational Factors Affecting Exercise Involvement: The Importance of Enjoyment', *Research Quarterly for Exercise and Sport* **56**(3), 275–282.
- Wilk, R. & Johnson, M. J. (2014), Usability feedback of patients and therapists on a conceptual mobile service robot for inpatient and home-based stroke rehabilitation, in '5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics', pp. 438–443.
- Williams, A., Stephens, R., McKnight, T. & Dodd, S. (1991), 'Factors affecting adherence of end-stage renal disease patients to an exercise programme.', *British Journal of Sports Medicine* **25**(2), 90–93.
- Wilson, E. J. & Sherrell, D. L. (1993), 'Source effects in communication and persuasion research: A meta-analysis of effect size', *Journal of the Academy of Marketing Science* **21**(2), 101.
- Winkle, K., Caleb-Solly, P., Turton, A. & Bremner, P. (2018), Social Robots for Engagement in Rehabilitative Therapies: Design Implications from a Study with Therapists, in 'Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction', HRI '18, ACM, New York, NY, USA, pp. 289–297.
- Winkle, K., Caleb-Solly, P., Turton, A. & Bremner, P. (2019), 'Mutual Shaping in the Design of Socially Assistive Robots: A Case Study on Social Robots for Therapy', *International Journal of Social Robotics* .
- Winkle, K., Lemaignan, S., Caleb-Solly, P., Leonards, U., Turton, A. & Bremner, P. (2019), Effective Persuasion Strategies for Socially Assistive Robots, in '2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)', pp. 277–285.
- Wu, Y.-H., Fassert, C. & Rigaud, A.-S. (2012), 'Designing robots for the elderly: Appearance issue and beyond', *Archives of Gerontology and Geriatrics* **54**(1), 121–126.
- You, S. & Robert Jr., L. P. (2018), Human-Robot Similarity and Willingness to Work with a Robotic Co-worker, in 'Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction', HRI '18, ACM, New York, NY, USA, pp. 251–260.