Report from Dagstuhl Seminar 19081

Verification and Synthesis of Human-Robot Interaction

Edited by

Rachid Alami[1], Kerstin I. Eder[2], Guy Hoffman[3], and Hadas Kress-Gazit[4]

1 **LAAS – Toulouse, FR,** `rachid.alami@laas.fr`
2 **University of Bristol, GB,** `kerstin.eder@bristol.ac.uk`
3 **Cornell University – Ithaca, US,** `hoffman@cornell.edu`
4 **Cornell University – Ithaca, US,** `hadaskg@cornell.edu`

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 19081 "Verification and Synthesis of Human-Robot Interaction". This seminar brought together researchers from two distinct communities – Formal Methods for Robotics, and Human-Robot Interaction – to discuss the path towards creating safe and verifiable autonomous systems that are compatible with humans.

Seminar February 17–22, 2019 – http://www.dagstuhl.de/19081

2012 ACM Subject Classification Computer systems organization → Robotic autonomy, Human-centered computing → Human computer interaction (HCI), Software and its engineering → Formal methods

Keywords and phrases Formal Methods, Human-Robot Interaction

Digital Object Identifier 10.4230/DagRep.9.2.91

Edited in cooperation with Ross A. Knepper

1 Executive Summary

Hadas Kress-Gazit (Cornell University – Ithaca, US)
Rachid Alami (LAAS – Toulouse, FR)
Kerstin I. Eder (University of Bristol, GB)
Guy Hoffman (Cornell University – Ithaca, US)

There is a growing trend in robotics moving from industrial robots that work physically separated from people to robots that collaborate and interact with people in the workplace and the home. The field of human-robot interaction (HRI) studies such interactions from the computational, design and social points of view. At the same time, there is growing interest in research regarding the safety, verification and automated synthesis of behaviors for robots and autonomous systems. The fields of formal methods and testing, which focus on verification and synthesis of systems, aim to model systems and define and prove specifications over these systems; in the context of robotics, these techniques take into account the robot dynamics and its interaction with its changing and uncertain environment.

However, a human collaborating with a robot is not just part of the robot's environment, but an autonomous agent with intentions, beliefs, and actions that mesh with those of the robotic agent. This raises new research questions related to verification and synthesis including what appropriate models for human-robot interaction would be; whether and how algorithms for HRI can enable verification; how to take the human into account in automatic synthesis of robotic systems; and what (if any) guarantees can be provided with a human in the loop.

To date, very little work has explored questions of verification, safety guarantees and automated synthesis in the context of Human-Robot Interaction. HRI has modeled humans computationally but not from a verification point of view and without providing guarantees. Furthermore, there are rarely any formal specifications in the computational HRI literature; validated objective metrics for evaluation are also scarce. The verification and synthesis community has mostly focused on the robot's autonomous behavior and its environment, and not paid much attention to the integral presence of the human or the interaction, including the psychological, social, and intentional aspects of human activity.

In this seminar we bring together experts in computational HRI, verification of autonomous systems, formal methods, and cognitive and social psychology to exchange ideas, define research directions, and foster collaborations toward a new theory and practice of verifiable HRI.

2 Table of Contents

Working groups

3 Overview of Talks

3.1 Education in HRI

Henny Admoni (Carnegie Mellon University – Pittsburgh, US)

This short talk provides an overview of the field of HRI for education. It covers research from this field broadly at an introductory level by providing examples of a selection of projects. There is an overview of the benefits, models, and challenges of HRI for education.

3.2 Computational approaches in HRI: Examples and Discussion

Rachid Alami (LAAS – Toulouse, FR) and Guy Hoffman (Cornell University – Ithaca, US)

Brief review of design objectives for the implementation of robots that engage in collaboration with human. Teamwork and joint action: collaboration and coordination. Presentation of a set of examples.

3.3 Human-Robot Handovers

Maya Cakmak (University of Washington – Seattle, US)

Handovers are an essential capability for personal robots that are intended to assist humans or collaborate with humans in different environments, such as homes, hospitals, factories. This talk characterizes the problem of human-robot handovers, identifying dimensions in which handovers can vary and ways in which we can measure the success and quality of a handover.

3.4 Two examples of how we used verification techniques in HRI

Kerstin I. Eder (University of Bristol, GB) and Dejanira Araiza-Illan (ARTC – Singapore, SG)

These slides show how we used a combination of different verification techniques to gain confidence in the correctness of autonomous systems that interact with humans, i.e. HRI. The first investigates the front-end of the system development process, where specifications are developed/translated into designs, e.g. in Simulink. The important question is whether these Simulink designs preserve the intent of the specification. This is so important because if the final coded system does not work, it is very useful to know whether this is because there is a coding bug, or whether there is a bug in the design. As such, verification should be done as early as possible, including in particular the design level. We propose the introduction of assertions, written in Simulink, and added directly to the Simulink model. Whether the design satisfies these assertions can then be determined using two techniques. Some assertions can be tested, i.e. those that are ground. Others, i.e. those that contain variables, can be checked using theorem proving, ideally automatic theorem proving techniques. In a different context, we developed a coverage-driven simulation-based verification technique (CDV) for a robot to human hand-over task. We present the different components of a CDV testbench, the System Under Test, i.e. the robotic code, the test generator, the checker and coverage collector. In particular, we show some of the requirements used to encode assertion monitors to flag issues during simulation as well as the coverage models we used. Associated papers are linked into the slides. Please contact us for further information.

3.5 A Simple Model

Rüdiger Ehlers (Clausthal-Zellefeld, DE)

During the seminar, we identified a lack of compatibility between the models of human behavior in the field of Human Robot Interaction (HRI) and the models used by the formal methods community. This short talk provided some initial ideas on how the latter type of model can be brought closer to the former one. One concrete Markov Decision Process (MDP) formulation of a scenario in which a human and a robot pass by each other in a narrow hallway was given. The MDP already captures *some* aspects from the talk given by Marc Hanheide (Section 3.6). Starting from this formulation, the idea is to add additional HRI aspects one by one. Along the way, the necessary concepts to increase the scalability of analyzing the MDPs of the particular resulting shape can be researched.

3.6 Social Navigation: Problem Statement and (Some) Solutions

Marc Hanheide (University of Lincoln, GB)

Social navigation constitutes an area of research in robot navigation, to explicitly respect social norms, to model human intent in navigation, and ultimately provide robots with the skill to safely and appropriately navigate in the presence of humans in a shared space. First, two main contexts are evident in literature: One is to avoid and circumvent humans in order to ensure safe passage and increase acceptability and perceived safety (trustworthiness) of robots for humans. Second, it studies suitable ways to approach humans to facilitate and support further interaction (e.g. dialogues, object hand-overs, and other forms of engagement). Literature reports on different approaches to accomplish these objectives, such as social cost- or utility maps, joined multi-agent planning explicitly modelling possible human behaviour in response to robots, and approaches employing computational models of games theory or explicit discrete state transitions to predict human-robot spatial interaction patterns. Some of these models lend themselves quite readily to further investigation in the context of form synthesis and verification.

3.7 HRI Specifications and Natural Language

Christoffer R. Heckman (University of Colorado – Boulder, US)

Formal human-robot interaction requires the creation of some specification for which there are many different forms. They might include specification of safety that consider some behaviors proximal to humans that can include restrictions of state space, or some more amorphous social contracts that cannot be broken. In this short talk, I consider how natural language grounded in the spatial environment might also be considered as a form of specification. On the one hand, this mode is natural and is the form in which we as humans find simplest to define instructions (e.g. "go over there and hang out for a bit"). On the other, natural language has an enormous prior information set from which it draws and is also variably dependent on the environment. I consider some techniques in machine vision and dynamical systems that have been in some sense brought to heel through the success and proliferation of data-driven techniques, but I also identified a few remaining challenges related to how one might define specifications in natural language. Finally, I give some considerations for how this might work in the future through joint vector embeddings and ontological grounding.

3.8 Iterative Design of Verifiable Human Models

Guy Hoffman (Cornell University – Ithaca, US)

I propose a human-centered process of iteratively designing formal models for human-robot interaction. The framework is inspired by iterative practices in human-centered design (HCD), often conceptualized as a cycle of four phases: observation, ideation, prototyping, and testing. I propose mapping these onto existing practices in formal methods in robotics, such as model-building, testing, and verification. In the proposed process, samples of human participants interact in human-participant studies with robots synthesized from intermediate models and specifications. Verification methods are used to test and update these models and specifications, leading to newly synthesized controllers that are then tested with an additional sample of human interactants. The proposed approach suggests a number of research questions, including how to optimally sample human interactants, and whether and how to update models from outcomes of the interaction studies.

3.9 Verification of Autonomous Robots, a Roboticist Bottom Up Approach

Felix Ingrand (LAAS – Toulouse, FR)

Complete validation and verification of the software of an autonomous robot (AR) is out of reach for now. Still, this should not prevent us from trying to verify some components and their integration. There are many approaches to consider for the V&V of AR software, e.g. write high level specifications and derive them in correct implementations, deploy and develop new or modified V&V formalisms to program robotics components, etc. We propose an approach which rely on an existing robotics specification/implementation framework (GenoM) to deploy robotics functional components, to which we harness existing well known formal V&V framework (UPPAAL, BIP, FIACRE/TINA). GenoM was originally developed by roboticists and software engineers, who wanted to clearly and precisely specify how a reusable, portable, middleware independent, functional component should be written and implemented. Many complex robotic experiments have been developed and deployed using GenoM and it is only recently that its designers realized that the rigorous specification, a clear semantic of the implementation and the template mechanism to synthesize code opens the door to automatic formal model synthesis and formal V&V (offline and online). This bottom up approach, which starts from components implementation, may be more modest than the top down ones which aim at a larger and mode global view of the problem. Yet, it gives encouraging results on real implementations on which one can build more complex high level properties to be then V&V.

3.10 Multimodal Dialog in HRI

Ross A. Knepper (Cornell University, US)

Multimodal dialog exploits natural human communicative abilities to mediate human interaction in service of a joint activity or shared task. Modalities include any combination of speech, gesture, facial expression, eye gaze, body language, and gross body motion. Dialog involves a back and forth exchange, in which previous communicative acts become context in which to interpret later ones. The multiple modalities within a single communicative act also serve as context for understanding one another, which provides redundancy and makes the problem more tractable. This presentation walks you through the problem statement for grounding multimodal acts in a symbolic basis.

3.11 Task-Agnostic HRI

Ross A. Knepper (Cornell University, US)

There is an unbounded quantity of HRI tasks and scenarios for which domain-specific details are needed to specify and model correct behavior. In contrast, there is a set of behaviors that transcend the details of task and focus on the establishment of maintenance of a cooperative team, consensus around goals and intentions within the task, and understanding of the partition between shared and individual decision making. By formalizing and verifying these behaviors, many different HRI tasks benefit indirectly.

3.12 Testing, Verification, Synthesis Tutorial

Hadas Kress-Gazit (Cornell University – Ithaca, US) and Kerstin I. Eder (University of Bristol, GB)

This tutorial provided a basic introduction to terminology used in the context of testing, verification and synthesis with example applications in HRI.

3.13 Toward a human model – assessing perception-action coupling

Ute Leonards (University of Bristol, GB) and Kerstin I. Eder (University of Bristol, GB)

A key challenge for certification, verification and validation of robots interacting with people outside factory settings is how to ensure safety for all possible situations the robot might encounter. Ultimately, human behaviour remains unpredictable as there are far too many ways people could potentially interact with robots over and above those intended and accounted for within the design process. Efforts to solve this challenge include the application of increasingly complex cognitive models in the robot to predict human behaviour, including such aspects as theory of mind and other psychological theories on human social interaction, complex learning rules and so on; models that loose flexibility and processing speed.

Instead, we propose to go back to a very basic model of human action prediction, essentially a type of sanity checks. The model is based on the observation that human beings are embodied, and every task (action) involves a motor response that is usually tightly coupled to sensory input and thus the environment it is performed in (see Gibson's affordance models, 1979). Such perception-action coupling, be it for gait, hand, eye movements or a combination of these, is predictive for any given individual. Any deviation from this basic perception-action coupling in human behaviour, e.g. to perform an action in a different way to that expected or to perform a different action, leads to a delay, i.e. noticeable hesitation, to account for the required decision making time under cognitive control and the change in motor planning. In other words, the sanity check here would be a time-critical but simple check of whether the predicted action is performed: any decision making for predefined actions occurs within a small, predefined time window and is restricted to a small number of movement alternatives. Any temporal delay means therefore that the person is most likely not performing the task the robot is expecting them to perform, or at least not in the way the robot is expecting them to (e.g. the person is distracted); and safety measures should be taken.

The clear advantage of such a basic model would be increased usability as the model is quite generic. Research into developing a basic, safety-focused model for a variety of application domains would provide us with new insights into the feasibility and limitations of this approach. Compared to the myriad of application-specific models of increasing complexity that are currently available or under development, this human-centric approach promises simpler models, flexibility of use and computational efficiency.

3.14 Robots in Therapy

Shelly Levy-Tzedek (Ben Gurion University – Beer Sheva, IL)

There is an unmet need in therapy for clinician hours. Specifically, in post-stroke rehabilitation, patients need to preform repetitive practice; but without an accompanying therapist, compliance is low. One way to fill this “care gap” is to enlist robots. There are physically assistive robots, which can help the patient perform the task by moving their impaired limb, and there are socially assistive robots, which can help people by motivating them to perform the exercise and giving them feedback. If a therapist is not present during the exercise, however, there is potential damage that can ensue. Thus, the robotic system should be able to model, and respond to, the person’s affect (e.g., work by Jamy Li), intent (e.g., work by Hennt Admoni), and motor performance (e.g., work by Shelly Levy-Tzedek).

3.15 Physical HRI

Todd Murphey (Northwestern University – Evanston, US)

This talk briefly described some of the needs and challenges associated with using robots to physically assist and train people. Several key points about how physical Human Robot Interaction differs from other kinds of HRI were made. These include that the person and robot interact with each other through forces, and these forces have both mechanical effects–e.g., they can stabilize and destabilize someone–and communication effects–they can help someone learn from physical interaction. Several examples of what verification might look like in the context of physical assistance and rehabilitation were discussed, including the need to keep someone safe while avoiding overassistance.

3.16 Formal Specification Patterns for Sanity Checking

Kristin Yvonne Rozier (Iowa State University, US)

As a community, we have identified choosing the right (safe, progressing) next-action as one of the biggest challenges in human-robot interaction; doing this sufficiently quickly adds to the challenge. Sanity checking provides a tractable answer to this challenge. We exemplify sanity checks over the mission of the ExoMars Schiaparelli Lander and then generalize the patterns we so often encode for autonomous spacecraft, aircraft, and robots. Sanity checks are unsatisfying and unintuitive: we are unaccustomed to encoding common sense in the form of requirements and unsatisfied by the lack of diagnosis they provide. The necessity of choosing a next-action in real time, within the limitations of embedded computation requires us to sacrifice the satisfaction of diagnosing why and how a particularly interesting and complex error occurred in favor of the just-in-time determination that we need to switch to "safe mode." Motivated by the goal of choosing correctly from among the small, finite set of possible next-actions any automated system can execute, sanity checks provide a promising way forward, that we can monitor and enforce on-board via runtime verification, e.g., with R2U2 (http://temporallogic.org/research/R2U2/).

3.17 Interactive Autonomy: a human-centered approach for safe interactions

Dorsa Sadigh (Stanford University, US)

Reward functions are formal specifications used to describe how a robot should act or interact with humans. Similar to specifications, coming up with reward functions can be challenging too. We would like to learn these specifications either from demonstrations or preferences. However, teleoperating robots with high degrees of freedom is quite challenging so learning reward functions from demonstrations can be limiting. Instead, we propose an approach to actively generate new scenarios and query humans in order to learn their preferences from a combination of pariwise comparisons and limited expert demonstrations.

4 Working groups

4.1 Physical Human-Robot Interaction Discussion Group

Brenna D. Argall, Kerstin I. Eder, Christopher R. Heckmann, Ute Leonards, Todd Murphey, Kristin Yvonne Rozier

This breakout group discussion focused on how and when verification techniques may be used to increase trustworthiness of a physical human-robot system. Trustworthiness is a mutual effect in physical Human-Robot Interaction (pHRI)–the human must trust the automation and the automation must trust the human (often with the automation regulating the exchange of decision authority). Parts of the system may be verifiable in a classical formal methods sense–for instance, the computing elements–but the group agreed that it seems unlikely that the human is a classically verifiable component. The discussion primarily built up a model of where and how verification methods can be applied. The resulting model included many interacting components: the human, the robot, the forceful interactions between them, the computed combined model of the human and robot, observations of the combined system, online and offline machine learning algorithms responsible for building models of the combined system, the computing elements, and physics/psychophysics that constrain possible behaviors. These interactions form a complex network of interdependent components, where each component of the network could be formally modeled and verified and the interactions between them could be modeled and verified. An important insight is that a person or multiple people can always choose to undermine a pHRI system, so guarantees of safety are necessarily limited to making sure that the rest of the automation is not responsible for failure.

4.2 An Iterative Workflow for HRI Model Repair

Frank Broz, Jan Kretinsky, Nils Jansen, Hadas Kress-Gazit, Guy Hoffman

In this breakout session, we discussed an iterative workflow enabling model repair for human-robot interaction tasks. Our approach is anchored in a probabilistic joint human-robot interaction model drawing on separate human and robot state transition graphs. The workflow includes learning initial parameters for a human model from human-human interaction datasets, and evaluating these vis-a-vis specifications drawn from the social psychology literature. A synthesized robot controller is then composed with the human model and put to test in a human participant experiment. The outcomes of this experiment is used to both study guarantees on the interaction and refine the model as part of the iterative improvement cycle. As technique for model refinement we plan to explore model repair under temporal logic constraints. We decided to begin by implementing this workflow on mutual gaze and handover tasks.

4.3 Synthesis-Aided End-User Programming of Interactive Robots

Maya Cakmak, Ivan Gavran, Jana Tumova, Aurelie Clodic, Shelly Levy-Tzedek, Laurel Riek, Hadas Kress-Gazit, Marc Hanheide, Rüdiger Ehlers

Programming interactive robots to create new applications is challenging, even for experienced software developers, due to the complexity of concurrently handling multiple input channels while generating actions across multiple modalities (speech, sounds, gesture, gaze, text on screen, facial expression, motion). Program synthesis can enable writing better programs in less time and with less prior expertise by automating parts of the programming process. However, it is currently unclear how what parts of the programming process can be automated, what user input can be captured and translated into specifications that can be used for synthesis, and what synthesis methods are appropriate. The goal of this project is to identify opportunities for applying synthesis to improve the process of programming interactive robots.

4.3.1 Specific Outcomes

- Choose domains/tasks focused on robots that socially interact with people, e.g. storytelling robot, language tutor robot, stress support robot
- Clarify what we mean by program/controller, e.g. finite state machines
- Identify prior formats of specifications, e.g. partial programs, correctness properties, interaction traces, sketches, etc
- Identify modalities for capturing user input, e.g. natural language, demonstration (in different ways), visual programming environments, text, etc
- Identify methods for translating user input to specs
- Identify methods for combining different combinations of specs to synthesize robot programs with certain properties
- Work out a running example for all of the above based on chosen tasks
- Make figures to communicate created knowledge/ideas
- Outline position/framing paper to communicate created knowledge/ideas
- Identify low-hanging novel research and discuss possible collaborative paper opportunities

4.4 A Framework for Synthesis-Aided End-User Programming of Interactive Robots

Maya Cakmak, Jana Tumova, Laurel Riek, Shelly Levy-Tzedek, Ivan Gavran, Aurelie Clodic, Hadas Kress-Gazit, Rüdiger Ehlers

Programming interactive robots to create new applications is challenging, even for experienced software developers, due to the complexity of concurrently handling multiple input channels while generating actions across multiple modalities (speech, sounds, gesture, gaze, text on screen, facial expression, motion). Program synthesis can enable writing better programs in less time and with less prior expertise by automating parts of the programming process. However, it is currently unclear how what parts of the programming process can be automated, what user input can be captured and translated into specifications that can be used for synthesis, and what synthesis methods are appropriate. The goal of this project is to identify opportunities for applying synthesis to improve the process of programming interactive robots.

4.5 HRI Skill: Social Navigation (Elaboration)

Marc Hanheide, Rüdiger Ehlers, Jana Tumova, Felix Ingrand, Kerstin I. Eder, Satoru Satake

This is an elaboration group from Tuesday with the goal of reaching more detailed and thought-through outcomes.

4.5.1 Specific Outcomes

- More detailed models, including a comparison between them
- Low hanging fruit for specific research projects
  - Simulation test generation
  - Basic proven runtime guards
- Technical and intellectual challenges
- Defining metrics for success
- Human models (approach, pass)
- Defining contexts of the area

4.6 Formalizing Flexible Collaboration in HRI

Ross A. Knepper, Morteza Lahijanian, Henny Admoni, Rachid Alami, Shanee Honig, Satoru Satake, Victor Fernández Castro

There is an unbounded quantity of HRI tasks and scenarios for which domain-specific details are needed to specify and model correct behavior. In contrast, there is a set of behaviors that transcend the details of task and focus on the establishment of maintenance of a cooperative team, consensus around goals and intentions within the task, and understanding of the partition between shared and individual decision making. By formalizing and verifying these behaviors, many different HRI tasks benefit indirectly.

4.6.1 Specific Outcomes

- Define a few simple, diverse exemplar tasks that we could implement
- Identify formalisms and models
- Provide a minimal necessary set of capabilities for a working model
- Identify a list of abstract capabilities that the system would ideally have
- Plan data collection study in each task

4.7 Model Repair for Models of Human

Jan Kretinsky, Frank Broz, Hadas Kress-Gazit, Guy Hoffman

Given a wrong model of a human a true property that does not hold in the model, how to fix the model to reflect it.

4.8 HRI Application Area: Healthcare and Therapy (Elaboration)

Jamy Jue Li, Erika Abraham, Victor Fernández Castro

This is an elaboration group from Tuesday with the goal of reaching more detailed and thought-through outcomes. Specific Outcomes:
- More detailed models, including a comparison between them
- Low-hanging fruit for specific research projects
- Technical and intellectual challenges
- Defining metrics for success
- Human models
- Defining contexts of the area

4.8.1 Rationale for HRI Therapy + Models

1. Human therapist will not remember all the details about the past because they have so many people they work with, thus a system model that tracks and adapts to many details that the human therapist simply cannot track over time will be helpful to use to adapt and personalize therapy to an individual in a way that a human cannot.
2. To maintain the therapist involvement and expertise, the therapist may be helped by receiving advice (i.e., recommendations) from the system based on information that the therapist may not remember or even perceive (such as minute but consistent changes in response times)
3. To be able to look at a wide variety of system parameters that could affect a metric of success (for example, the user's improvement in performance in an educational test), the model of the system could test many free parameters of the system that have been identified by the researcher and refined through discussion with the therapist

4.8.2 Overall Process

- *Example application overview.*
  Robot playing an activity with an autistic child where an adult therapist is leading the interaction. The adult therapist presses buttons on their tablet to initiate one of multiple activities for the robot to play, and to initiate specific actions within those games like the robot displaying a "happy" face, "sad" face, etc. in the order the therapist chooses. The child can also initiate actions on their tablet (such as them choosing for the robot to make a happy face) and also respond to the robot's questions using the tablet.

- *Data collection of interactions of users with the system*
  An example of data collection in HRI studies is 50 participants are run and interactions are videotaped, then annotated by 3 trained coders for participant affect (emotion), engagement and performance in a therapeutic task. The annotation is created by coders who view a 15 minute video clip of a child interacting with the robot and code at each 5 second time point the child's affect and the child's engagement. This timeline of codes will be matched with the time-stamped record of the system's log, which contains both the timing of the robot's actions and the timing of the human therapist's initiated actions for the robot.
- *Data collection is needed*
  What is annotated in the videos is a key consideration for the model, because it may be valuable to annotate a lot of additional variables (preferably automated annotation) that do not directly have a hypothesis around those variables because a model may be able to find a possible new relationship in the data.
- *Which types of model to use?*
  - A model for prediction is useful in this application (synthesizing a controller that is most likely to lead to success metric in the model)
  - A model for verification of the controller doesn't make much sense because it is hard to verify a human's behavior (however, it may be able to verify a limited subset of the human's behavior defined as the person's inputs with the system, which could be active input like tablet presses or passive input like smiles or eye gaze)
- *How to learn the model?*
  - Teaching-based isn't possible
  - Adaptation could be possible
- *How should the robot behave?*
  Personalization: Coupling of either two robot responses (robot responds with both a smile "you're correct" and a visual light) or two robot stimuli/prompts (delivery of both robot emotion, e.g., happy face, and another perceptual cue in tandem)

4.8.3 Operational Ideas

- What to do with the model?
  - Option 1: Find patterns in data executions In a large dataset of interactions that have been annotated, it could be possible to determine patterns in the data leading to insights of the relationship between the robot's and human's indicators/properties and the desired goal metric
  - Option 2: Give advice to therapist. In many situations where the therapist is involved in the therapy process with the technical system and their expertise is used to monitor the user's behavior and help the user, giving the therapist advice rather than autonomous adaptation (i.e., the system simply deciding what action to take) is a more acceptable strategy that respects the preference for the human therapist to always make the final call. Whether the human therapist follows the advice or not is a variable that the system can track. Advice-giving for therapists requires some level of explanation to the therapist about why each advised option is being presented in order to increase transparency to the therapist, which still being minimal cognitive load to distract from the therapist's main task of assisting the user. One way in which a graphical interface could present this to the therapist is by explaining what purpose (i.e., model strategy) each advice action corresponds to, such as: 1) system judges the success chance is highest with advised action #1; 2) system picks action #2 to improve ambiguity

in the model by testing this action; or 3) system judges action #3 will led to high success change for alternative outcome that could be desired (such as to increase child happiness) Types of advice: Parameters of therapy vs. robot behavior parameters
- Option 3: Make predictions about child behavior. In a large dataset of interactions that have been annotated, it could be possible to identify precursors to success metrics (e.g., child presses on tablet within 5 seconds) or child behaviors (e.g., child stands up and leaves) using characteristics captured by the system.
- Option 4: Optimize model learning to improve the model. In the event that the model cannot disambiguate between how two paths in the model affect the success metric, the model can be used to generate multiple test cases corresponding to each of the paths to the model. The resulting group of executions could then be used to refine the model to determine which of the paths leads to a better success metric result.

Model construction.

Variables to capture
- Order of the trials in a task (e.g., order of the emotions that the game goes through)
- Collect the parameters of robot appearance and behavior that can be varied, then run these by the therapist
- Usability test with some of the parameters
- Design of executions
- Vary parameters in multiple executions
- For example, using multiple variable time delays in an interaction with a robot between [0,25 s – 4 s] at each verbal response of the robot can provide a dataset of variable time delays to try to identify which time delay results in the best human performance (or other measure) of the system
- Another example is using multiple embodiments in the learning activity, such as a photo of a robot, a screen image of a robot, an actual robot

4.8.4 Other Considerations

Time needed to be interdisciplinary
Takes time to model data in the correct form, to collect the data, how to present the advice to the therapist.

4.9 HRI Application Area: Healthcare and Therapy

Jamy Jue Li, Erika Abraham, Victor Fernández Castro, Shelly Levy-Tzedek

In a scenario where a human therapist uses a robot to monitor and help a user (for example, in physical therapy or social skills therapy), a model could provide advice to the therapist on how the robot should behave while explaining the strategy for the advice (for example, highest predicted primary outcome, highest predicted secondary outcome or disambiguating how two potential paths in the model affect outcomes). User parameters to be collected for prediction could include order of task trials in an activity, type of trials (e.g., reach task vs push task; emotion task vs informational task), and design characteristics of the robot.

Design executions could also be autonomously generated to explore how new robot parameters (for example, latency in robot response) affect outcomes. These may be personalised per individual. A key challenge may be the time needed to construct the model, collect data and design how to present advice to the therapist.

4.10 HRI Application Area: Industrial Assembly (Elaboration)

Björn Matthias, Michael Gienger, Daniele Magazzeni, Alessandro Cimatti, Dejanira Araiza-Illan

This is an elaboration group from Tuesday with the goal of reaching more detailed and thought-through outcomes.

Specific Outcomes

- More detailed models, including a comparison between them
- Low-hanging fruit for specific research projects
- Technical and intellectual challenges
- Defining metrics for success
- Human models
- Defining contexts of the area

4.11 Verification Modeling for Physical HRI

Todd Murphey, Kristin Yvonne Rozier, Kerstin I. Eder, Brenna D. Argall, Ute Leonards, Christopher R. Heckman

This is a continuation of the Tuesday group, focusing on how and when verification techniques may be used to increase trustworthiness of a physical human-robot system. Trustworthiness is a mutual effect in physical HRI—the human must trust the automation and the automation must trust the human (often with the automation regulating the exchange of decision authority). Parts of the system may be verifiable in a classical formal methods sense, but it at least seems unlikely that the human is a classically verifiable component. Nevertheless, in the context of an otherwise formally understood system, the human model could be falsified. The purpose of this discussion will be to build up a model of where and how verification methods can be applied.

Specific Outcomes

- Data-driven modeling and its properties
- Learning and active learning of specification semantics from continuous time/space physical HRI
- Which pieces of a physical HRI system can be formally verified using theory?

- Among the pieces that can be theoretically verified based on models, how can those verified pieces be composed?
- What are a list of pHRI applications that are both important and have reasonable decompositions into analyzable pieces? Semi-rigid exoskeletons are probably an example, but soft-body exoskeletons may not be. Understanding what constitutes the division would be helpful.
- What is at least one concrete project that could be accomplished in a 2-3 year period?
- Diagram describing key needs and challenges

Participants

- Erika Abraham
RWTH Aachen, DE
- Henny Admoni
Carnegie Mellon University – Pittsburgh, US
- Rachid Alami
LAAS – Toulouse, FR
- Dejanira Araiza-Illan
ARTC – Singapore, SG
- Brenna D. Argall
Northwestern University – Evanston, US
- Frank Broz
Heriot-Watt University – Edinburgh, GB
- Maya Cakmak
University of Washington – Seattle, US
- Alessandro Cimatti
Bruno Kessler Foundation – Povo, IT
- Aurelie Clodic
LAAS – Toulouse, FR
- Kerstin I. Eder
University of Bristol, GB
- Rüdiger Ehlers
Clausthal-Zellefeld, DE
- Victor Fernandez Castro
ENS – Paris, FR
- Ivan Gavran
MPI-SWS – Kaiserslautern, DE
- Michael Gienger
Honda Research Europe – Offenbach, DE
- Marc Hanheide
University of Lincoln, GB
- Christoffer R, Heckman
University of Colorado – Boulder, US
- Guy Hoffman
Cornell University – Ithaca, US
- Shanee Honig
Ben Gurion University – Beer Sheva, IL
- Felix Ingrand
LAAS – Toulouse, FR
- Nils Jansen
Radboud University Nijmegen, NL
- Ross A. Knepper
Cornell University, US
- Hadas Kress-Gazit
Cornell University – Ithaca, US
- Jan Kretinsky
TU München, DE
- Morteza Lahijanian
University of Colorado – Boulder, US
- Ute Leonards
University of Bristol, GB
- Shelly Levy-Tzedek
Ben Gurion University – Beer Sheva, IL
- Jamy Jue Li
University of Twente, NL
- Daniele Magazzeni
King's College London, GB
- Björn Matthias
ABB AG Forschungszentrum – Ladenburg, DE
- Todd Murphey
Northwestern University – Evanston, US
- Laurel Riek
University of California – San Diego, US
- Kristin Yvonne Rozier
Iowa State University, US
- Dorsa Sadigh
Stanford University, US
- Maha Salem
Google UK – London, GB
- Satoru Satake
ATR – Kyoto, JP
- Jana Tumova
KTH Royal Institute of Technology – Stockholm, SE