Reducing Cognitive Load in HRI: A Proposed Research Strategy for the RAIL Lab

Sarah Storer, Sonia Chernova, Tesca Fitzgerald

How can social robotics help reduce cognitive load when humans and robots work together?

"The Robot Autonomy and Interactive Learning (RAIL) research lab focuses on the development of robotic systems that operate effectively in complex human environments, adapt to user preferences and learn from user input."

Introduction

If researchers want to incorporate robots into daily life, they need to learn how to make it easy to interact with them. There can, however, be many different definitions of "easy to interact with". In the pursuit of finding something that works for everyone, we can look to psychologists who study how the human brain learns, thinks, and solves problems. There is a branch of psychology called cognitive load theory (CLT). Cognitive theorists focus on the existing cognitive structures in humans to design systems and instructions that make the best use of their resources. Cognitive load is defined as the load that occurs in working memory during learning. By reducing the waste of mental resources in human-robot interaction, we are tailoring the way robots interact to the way humans are made to be interacted with.

There can be huge consequences when cognitive load is passed over while designing a system. In his piece, *The Design of Everday Things* [2], Donald Norman states that "users shouldn't need an engineering degree to figure out what a device does." Unnecessary complication leads to systems and devices that become easily frustrating to use as well as especially difficult for people with mental and physical disabilities. These results often lead to failed projects and failing companies.

The fields of social robotics and cognitive load theory have seldom if ever collided. In one study [3] that completed a literature review of CLT and HCI fields, 0.003% of papers in ''Guide to Computing Literature", a database provided by the Association of Computing Machinery (ACM), included something about CLT in their title or abstract. Those 65 or so papers have touched on fields of HCI like multi-modal interfaces, game design, and decision making, but nothing approached the field of social robotics. This study aims to apply what psychologists have learned about CLT to help researchers develop robots that communicate their intentions in a way that is easy for humans to understand. Our goal is to define cognitive load and lay out a road map for pursuing this research topic.

Cognitive Load

A 2010 article from the journal *Computers in Human Behavior* describes the back and forth nature of defining the specifics of cognitive load over the last few decades. The work of J. Sweller explains that CLT has three key pieces: intrinsic load, extraneous load, and germane load. Intrinsic load is defined as the effort required of the brain because of the latent complexity of the task. Extraneous load is caused by the inappropriate presentation of material or requiring the performance of actions that are irrelevant to learning. Germane load is generally seen as beneficial for learning where the brain is working on building schema in long term memory. The three types of load have been theorized to be additive or at the very least connected in some way. [3] This fact makes it exponentially more difficult to test individual pieces of CLT in a lab setting.

There are some guidelines posited by psychologists to adjust the cognitive load in certain learning circumstances. It is thought that intrinsic load is immutable. For this reason, researchers focus more on minimizing extraneous load and maximizing germane load. The three key methods of doing so, split-attention, modality, and redundancy are discussed in more detail in an article by Hollander, Nina et. al. [3] Future researchers should apply these guidelines to HRI situations to see if we can show the importance of legible motion and intention from a psychological perspective.

Materials and Methods

Evaluating cognitive load in a lab setting for HRI is mainly done using the NASA Task Load Index (NASA-TLX). The NASA-TLX was originally developed as a paper and pencil questionnaire by NASA Ames Research Center's Sandra Hart in the 1980s. The NASA-TLX has become the gold standard for measuring subjective workload across a wide range of applications. It is a two-part survey that measures cognitive load in 6 parts: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. These six 0-100 scores are called the "raw TLX". The second part takes each of the six cognitive load factors and asks the participant to rank the importance of each factor in pairwise comparisons. Many researchers however only consider the "raw TLX" because it has been proven that administering only the first part of the survey increases experimental validity. [4] Since we want to minimize cognitive load in an HRI scenario, the NASA-TLX fits our needs almost perfectly. We can use this multidimensional self-reporting to our advantage as a means of quantifying something that is often very difficult to put into words.

The NASA-TLX makes measuring overall cognitive load easier but, as mentioned earlier, it is difficult to split apart the influences of intrinsic, extraneous, and germane cognitive load. No single survey may be able to complete that goal. However, we suggest that researchers create a second survey to measure more attitudinal variables than the NASA-TLX such as perceived risk and “weirdness” of the robots actions. Taking these into account will allow researchers to see if there is more going on than just what the NASA-TLX can show us.

Our lab’s chosen robot for this study is the Fetch Mobile Manipulator, the flagship robot from Fetch Robotics. This robot has many advantages besides being made specifically for HRI research. The Fetch comes with an arm with 7 degrees of freedom and an attached modular gripper that can be used to interact with objects in the real world. This robot also comes equipped with a speaker to simulate speech, a swiveling head with two types of cameras (one 2D and one 3D), and 2 IMUs (one in the base and one in the gripper). All of these stock sensors make it easy to interface with human participants using RVIZ, Gazebo, and ROS integration. [5]

Research Questions	Sampling & Instruments	Participants	Data Analysis	Hypothesis
RQ1: What, if any, cognitive load exists when a robot is teaching a simple task?	NASA TLX Survey after every 3 tasks, Fetch Robot, Colored Blocks, 25 trials, Video, Capture Timing	25 Participants	Correlate TLX scores and attitudinal survey along with timing information from the trial video.	Intrisic cognitive load will exist in any HRI task. ------------------------- Extraneous cognitive load will be high. ------------------------- Germane cognitive load will be present but difficult to capture.
RQ2: Can implementing legible motion decrease extraneous cognitive load in the task learning?	Split Attention, Modality, Redundancy, NASA TLX Survey, Fetch Robot, Colored Blocks, 25 trials with 3 parts each	25 Different Participants than RQ1 study	Correlate TLX scores with timing information from the trial video and LM techniques used.	Moving in a way that humans expect will decrease extraneous cognitive load and make it easier to learn.

Figure 1. Guzdial Chart for Reducing Cognitive Load in HRI

Experiment Design

Designing an experiment for a field of research that hasn't been touched is one of the biggest challenges one can undertake as an undergraduate researcher. Thankfully, there are tools that exist to help organize this decision making process. Mark Guzdial, a professor at the Georgia Institute of Technology and the director of the College of Computing's Contextualized Support for Learning Lab, developed a modification of the Blumenfeld Chart for helping students plan their research and make goals concrete. The Guzdial Chart breaks down research questions into individual studies (one in each row), each with their own data collected, participants, data analysis, and hypotheses.

As shown in the Guzdial Chart, the first experiment will measure nascent cognitive load in a simple human-robot interaction by showing either normal or abnormal task-relevant motion. Our hypothesis for this experiment is that when the robot does something that the human participant considers strange there will be some amount of cognitive load present.

First, the robot's gestures to something simple, three different color blocks or numbers on a line, must be shown to participants and ranked on a scale from 1 to 10, normal to abnormal. These gestures must also be measured using a rating of subjective and attitudinal variables that the researcher should develop. These gestures should be created to be progressively irregular; changing variables such as timing, smoothness, joint movement, and distance. It is important to remember that humans tend to minimize distal angle space when gesturing towards an object. Once enough gestures have been rated and ranked, the researcher should group them into the different kinds of typical and atypical gestures that appear based on the surveys. In the first study, the participants will see the robot perform the simple gesturing task with 3 gestures from each group of gestures determined. The human participant will then repeat the pattern as a post-test to show that they have learned it from the robot teacher. After these three interactions, the participant will fill out the NASA TLX and a survey that measures attitudinal variables. The experiment should be videoed to allow the researcher to attempt to correlate response time with cognitive load. Using this information, the data should be analyzed to begin to draw conclusions about abnormal robot motion and cognitive load.

It should be mentioned that this experiment may prove either that there is cognitive load present or that there isn't. If there is no measurable cognitive load in this simple human-robot interaction, the researcher should consider different mechanisms that could be at play. If there is a measurable difference in cognitive load between normal robotic movements and atypical robot movements then the researcher should continue on to determine how legible motion can minimize extraneous cognitive load in human-robot interactions.

Discussion

Success for this project is the exploration of the confluence of HCI and CLT as well as laying the groundwork for future researchers who want to tackle this largely untouched field. We started out with a plan to In this sense, we have met all of our goals and left the research community in a better place than we found it. This process began approximately three years ago with a 7 DoF arm lovingly named Prentice. Since then, we have iterated over three platforms, explored multiple research areas, and re-did many hours of bring-up. Once we settled on the Fetch platform, narrowing down a research question was incredibly difficult.

Discovering the CLT space was interesting since the RAIL lab doesn't do much work with psychology. It took many discussions and many paper readings to understand the boundaries and see what research tools could be used to discover more. Since we confirmed HCI and CLT to be an interesting space and we have prescribed research methodologies discuss the next step, we would like to encourage some future researcher to attempt our plan and make their own changes.

Acknowledgments

First of all, I'd like to thank Sonia Chernova, my research advisor. She has guided me over the last three years through the ups and downs, starts and stops of this whole project. I also owe a lot to Tesca Fitzgerald, my graduate student mentor, for helping me formulate goals and milestones in the midst of the toughest parts of this project. Thank you to Bruce Walker and Tobias Wilson-Bates. Without their kindness, I definitely would not have graduated on time. The world needs more people like them.

Resources

[1] Gockley, Rachel, et al. "Designing robots for long-term social interaction." Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on. IEEE, 2005.

[2] D. Norman, The Psychology of Everyday Things. Basic Books, New York, 1988. In paperback as The Design of Everyday Things. Doubleday, New York, 1990.

[3] Hollender, Nina et al. “Integrating Cognitive Load Theory and Concepts of Human-Computer Interaction.” Computers in Human Behavior 26.6 (2010): 1278–1288. Computers in Human Behavior. Web.

[4] “Measurement Invariance of the Nasa TLX.” *SAGE Journals*, journals.sagepub.com/doi/10.1177/154193120805201946.

[5] Wise, Melonee, et al. “Fetch & Freight: Standard Platforms for Service Robot Applications”. Fetch Robotics Inc. San Jose CA, USA, fetchrobotics.com/wp-content/uploads/2018/04/Fetch-and-Freight-Workshop-Paper.pdf.