STATEMENT OF RESEARCH OVERVIEW, EXPERIENCE, AND DIRECTION

Alan Wagner
Georgia Institute of Technology

1. RESEARCH OVERVIEW

Some robots need to be social. The robots that take care of you, that protect you or that watch over your children should tailor their behavior to your needs as a person. These systems should relate to you and you should be able to relate to them. Moreover, they will need to utilize their social experience to guide and inform their behavior.

The challenge of developing a formal, computational means for representing the needs, motives, and behavior of a person is central to human-robot interaction. **My research draws on theories from game theory and psychology to create robots that 1) understand how people interact and 2) interact like a person.** My goal is to develop robots that can interact, learn from, and collaborate with ordinary people in a wide variety of social contexts. There are a number of factors that make this a difficult problem. First, human behavior varies tremendously from person-to-person, and often changes over time for the same person. Moreover, in addition to the robot's behavior, the context often influences a person's behavior in subtle and difficult to predict ways. Finally, humans can be strategic or deceptive in their behavior selection, generating behaviors that may not match their stated goals or motives.

In order to manage these challenges my work relies on a unique combination of game theory and psychological theory to formalize, capture, and understand the nuances that make human decision-making human. I borrow heavily from social psychology, behavioral economics, and artificial intelligence focusing on higher, cognitive, aspects of human-robot socialization such as relationship development, modeling of one's interactive partner, and reasoning about trust and deception. I utilize theories and methods from these fields to create robots that are capable of social interaction with an ordinary person in variety of different environments. The core of my work has focused on the development of the Interdependence Framework for Social Action Selection (IFSAS) that allows a robot to create and use game-theoretic representations to make predictions about its social interactions with a person and to use tools from interdependence theory to guide the robot's social behavior [1, 2].

My research is strongly interdisciplinary. I have collaborated with ethicists, doctors, and philosophers to understand the ramifications of creating robots that interact with people and to define prosocial use cases for the technology I develop [3, 4]. I use approaches from multiple different areas and strive to build bridges between the robotics community and other disciplines.

My research has been sponsored by grants from the Office of Naval Research, Naval Surface Warfare Center as well as by an Air Force Office Sponsored Research Young Investigator Award. The funding from these agencies has supported one graduate student which will complete his doctoral degree in December 2015, one graduated master's student, and two master's students which will graduate in May 2016.

The research I have completed to date has demonstrated that robots can be used to investigate a variety of social phenomena such as trust, deception, and stereotyping. My current and future research agenda addresses important problems such as robot led emergency evacuation and human-robot trust, the solutions to which will broadly impact society and transform the state of the art in robotics and other fields.

2. Research Experience

Human-Robot Trust. Beginning with my dissertation [5], an important focus of my research been to use IFSAS to develop methods that will allow a robot to both evaluate and respond to trust. I received an Air Force Young Investigator Program (YIP) award to explore the issues that surround human-robot trust and to begin to develop methods for trust repair.

Figure 1. An emergency evacuation robot signals for the person to go to the right.

The series of experiments funded by this award involved over 2100 human participants which interacted with simulated and real robots in an emergency guidance scenario which tasked the subjects with finding an exit in a maze-like office environment. Participants had the option of using a mobile robot to assist them in finding an exit or could search on their own.

Using this scenario we have shown that people tend to initially trust robot [6], trust drops precipitately after a single mistake [7], but can be repaired using repair methods such as promises and apologies if the repair is timed correctly [8]. In a live experiment (Figure 1) a robot lead people to a room, making mistakes along the way. The hallway was then filled with realistic smoke intentionally causing a fire alarm to sound. We then observed whether or not people followed the robot to an exit. To our surprise, 41 out of 46 participants followed the robots directions regardless of the robot's prior reliability. These results highlight the possible pervasiveness of overtust.

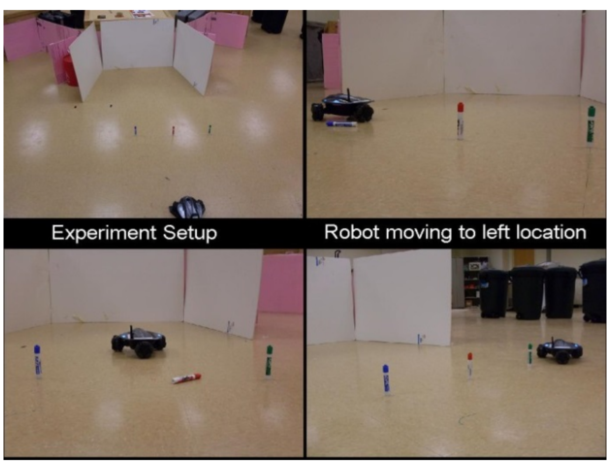

Figure 2. The robot chooses a location to hide in. It then feigns moving to a different location in order to knock over markers which will deceive the mark as to its trust location.

Robot Deception. In addition to investigating human-robot trust, my research has also used IFSAS to examine deception. I used an established definition for deception to examine which portions of the interdependence space warrant the use of deception. The interdependence space is a space capturing the relationship of different outcome matrices (computational representations of interaction) to one another. My analysis showed that the precusors for deception are dependence and conflict. Using this information, I developed an algorithm which mapped a robot's interactions in outcome matrix format to a portion of the interdependence space. Matrices located in a particular region of the interdependence space indicate a social situation which warrants the use of deception by the robot or the person. This algorithm was tested in simulation and later on a robot playing hide-and-seek [9].

In follow-up research I developed a method that allowed a robot to learn and use a model of its partner to determine which deceptive action would be the most efficacious [10]. This research was tested in a multi-robot hide-and-seek paradigm (Figure 2) in which a robot choose which of three different locations to hide in. Along the way to the site, the robot would knock over markers leaving a sort of trail indicating its decision. The results demonstrated that the robot could learn to generate false trails which influenced the robot being deceived to search the wrong location.

Stereotypes. In order for a robot to decide who to trust or who to deceive the robot must have a model of its interactive partner [11]. Yet, building a model from successive interactions with a person is impractical given the variety of social actions people typically employ [2]. I have developed a method which allows a robot to create and use categorical models of people to generate initial hypotheses about a newly encountered person based only on a person's perceptual features [12]. These stereotypes are learned from the robot's experiences with different types of people and used to construct a mapping from the partner's perceptual features to a learned category of partner model.

Figure 3. An experiment demonstrating robot stereotype learning and usage.

We have empirically shown that our method can be used by a robot to learn the tool preferences of different types of emergency responders (firefighters, police officers, and EMTs). The resulting stereotypes can then be used to predict which tool each type of responder (Figure 3) will need [13]. In research funded by ONR we extended these ideas to learning models of different categories of social environments such as a school, restaurant, and nursing home [14]. These situation-specific stereotypes allow the robot to predict the type of people it will encounter in an environment as well the actions that they will perform.

Deep Learning for Social Perception. Motivated by our stereotyping research, in joint work with Zsolt Kira, we sought to create categorical models of different environments from first-person video. We developed a computational process that utilizes the output maps generated by a multi-layer convolutional neural networks (CNNs) pre-trained on ImageNet to categorize scenes from blurry video taken in natural environments [15]. Convolutional neural networks learn a hierarchy of visual features ranging from low-level filters to object parts to entire objects themselves. The output maps generated by the fifth layer of a CNN correspond to higher-level objects and tend to display greater visual invariance and robustness than lower level features. Our results show that the process can be used to automatically divide video into self-similar segments. These segments correspond to distinct portions of one's time and place in an environment. For example, eating at a fast-food restaurant creates segments focused on entering the restaurant, ordering, sitting at a table, and leaving. We have shown that the initial segments experienced in an environment can be matched to previously experienced environments. One can then use the previously encountered environment to make predictions about the objects or segments that will be encountered in the near future.

Funding from ONR and Georgia Tech's Institute of Robotics and Intelligent machines has allowed us to create and demonstrate a working prototype of the system. The prototype uses the video feed from a camera to match a person's current environment to a previously encountered environment and then outputs the name of the type of environment. I believe that this work will serve as a conceptual and computational bridge between low-level sensor-based scene categorization techniques and high-level methods influenced by insights from human thinking.

My overarching hope is to connect newer insights from deep learning to higher-level methods from Artificial Intelligence.

3. Research Direction

My previous and current research demonstrates the potential value of combining game theory and psychological theory to guide and control a robot's social behavior. I intend to maximize the impact of this approach by simultaneously investigating the theoretical underpinnings for human-robot interaction while also pursuing applications in search and rescue and healthcare robotics.

Theory—From Perception to Planning, a Pipeline for Socially Interactive Robots. Ultimately my research agenda is directed towards the creation of an architecture that allows a robot to model its interactive partner from perceptual information and to use this information as well as previous social experiences to guide the robot's social actions and plans. My work will continue to investigate methods that integrate higher level social phenomena such as trust, deception, and social norms into this architecture. As a window to better understand creativity, I hope to investigate and develop computational methods by which a robot learns how to cheat while playing a game with person, for example. My recent work involving deep learning may serve as a computational bridge connecting human-robot interaction to computer vision. We have shown that the output maps generated by a Convolutional Neural Network can be used to identify and make predictions about different types of social contexts. In the future, I intend to show that these representations can also be used to identify and make predictions about different types of interactive partners.

Applications—Search and Rescue and Healthcare Robotics. I plan to continue investigating the issues that surround robot guidance during emergencies. As buildings get taller and crowds get larger, the challenge of quickly evacuating people during an emergency becomes even more critical. A robot which guides evacuees away from bottlenecks and towards open exits could save lives. Moreover, a robot which has the capacity to match different types of search and rescue personnel to their role during an emergency could quickly provide rescuers with the important information they need.

With respect to healthcare robotics, I am currently pursuing research funding that would use our method for scene categorization to provide situation awareness to visually impaired individuals. A prototype of the system that we developed demonstrated that we could identify different types of environments using live, first-person video within approximately 10 seconds of entering an environment. The system was able to identify and state the type of environment over headphones to a user. I believe that the system could be used to provide a person with the locations of the items on a store shelf, for example, or the seats in a restaurant.

Impact and Funding. My research agenda investigates problems that impact and inspire people. My work in deception, for example, generated over 70 different articles in the print, radio, and online media. The work was selected as Time magazine's #13 top inventions of 2010. In addition to inspiring people, my research tackles problems with significant potential to benefit both scientific community and society at large. For instance, a proposal that I am currently leading with several Georgia Tech faculty members and a physician form Children's of Atlanta, which is under consideration by the NSF, would examine and model the risks that result when a disabled child or teenager relies on an exoskeleton for mobility. This information would be used to develop technologies which use the child's behavior to predict potentially harmful situations and actively prevents them. As part of this research, we will also examine the ethical ramifications of creating a system which monitors user's behavior and has the power to prevent certain types of behavior with the potential to harm.

My position as a senior research scientist has allowed me to cultivate relationships with a number of program managers from various government agencies. I am currently pursuing support from AFOSR, ONR, NSF, AFRL, and the National Eye Institute to fund these research ideas. As a tenure-track professor obtaining a CAREER award will be a priority for me. I have participated in and assisted in the development of collaborative MURI proposals, a proposal for an NSF Engineering Research Centers (ERC), and an industry lead indefinite delivery/indefinite quantity (IDIQ) contract.

Most current robots do not adapt as well or as much to the person using the system as they should. My research strives to discover transformative methods which will a robot to relate to the person using the robot.

REFERENCES

[1] A. R. Wagner and R. Arkin, "Recognizing Situations that Demand Trust," in *20th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2011)*, Atlanta, GA, 2011.

[2] A. R. Wagner, "Creating and Using Matrix Representations of Social Interaction," in *Human-Robot Interaction (HRI)*, San Diego, CA, 2009a.

[3] R. C. Arkin, P. Ulam and A. R. Wagner, "Moral Decision-making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust and Deception," *Proceedings of the IEEE*, vol. 100, no. 3, p. 571 – 589, 2012.

[4] W. Barnett, A. R. Wagner and K. Keeling, "Social Robots and Older Adults: Some Ethical Concerns for Researchers," in *Proceedings of Fifth International Conference on Internet Technologies and Applications (ITA 13)*, Wrexham, North Wales, UK, 2013.

[5] A. R. Wagner, The Role of Trust and Relationships in Human-Robot Social Interaction, (Doctoral Dissertation). School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, 2009.

[6] A. R. Wagner and P. Robinette, "Towards Robots that Trust: Human Subject Validation of the Situational Conditions for Trust," *Interaction Studies*, vol. 16, no. 1, pp. 89-117, 2015 .

[7] P. Robinette, A. R. Wagner and A. M. Howard, "The effect of robot performance on human-robot trust," *IEEE Transactions on Human-Machine Systems (under review).*

[8] P. Robinette, A. Howard and A. R. Wagner, "Timing is Key For Robot Trust Repair," in *Seventh International Conference on Social Robotics (ICSR 2015)*, Paris, France, 2015.

[9] A. R. Wagner and R. C. Arkin, "Robot Deception: Recognizing when a Robot Should Deceive," in *Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA 2009)*, Daejeon, Korea, 2009.

[10] A. R. Wagner and R. C. Arkin, "Robot Deception: Providing Robots with the Capacity for Deception," *International Journal of Social Robotics*, vol. 3, no. 1, pp. 5-26, 2010.

[11] A. R. Wagner, "Developing Robots that Recognize when they are being Trusted," Palo Alto, 2013.

[12] A. R. Wagner, "Robots that Stereotype: Creating and Using Categories of People for Human-Robot Interaction," *Journal of Human-Robot Interaction*, vol. 4, no. 2, pp. 97-124.

[13] A. R. Wagner, "Using Cluster-based Stereotyping to Foster Human-Robot Cooperation," in *Proceedings of IEEE International Conference on Intelligent Robots and Systems (IROS 2012*, Villamura, Portugal, 2012.

[14] A. R. Wagner and J. Doshi, "Who, how, where: Using Exemplars to Learn Social Concepts," in *Proceedings of the International Conference on Social Robotics (ICSR 13)*, Bristol, UK, 2013.

[15] J. Doshi, Z. Kira and A. R. Wagner, "From Deep Learning to Episodic Memories: Creating Categories of Visual Experiences," in *Proceedings from the Third Annual Conference on Advances in Cognitive Systems (ACS 2015)*, Atlanta, GA, 2015.