Title: Towards Robotic Trust: A Measure of Human Capability in Human-Robot Teams

Introduction: From intelligent transportation vehicles to ordnance disposal robots to exoskeletons, service members depend on their robotic partners to better serve our country. However, robotic partners require consistently reliable control signals to perform optimally. People who are influenced by changing health or environmental factors are not consistently reliable when operating complex machines. As robots become ubiquitous throughout the military, the need for a robotic partner to both recognize and adapt to their human partner's changing abilities becomes critical, especially when operating in dynamic, complex, and uncertain environments. This need grows exponentially when a human-robot team faces an adversarial situation and a robotic partner's ability to reason about a human's cogency may shift the outcome of a life-or-death scenario.

This research is a stepping stone towards developing robots that can differentiate useful human commands from those that are dangerous or noisy due to the individual's unique control capabilities. When a human is distracted or fatigued, their ability to control their robotic partner is diminished[1] increasing the potential for undesired outcomes. Without a measure of *robotic trust,* (i.e. the robot's belief in the partner's capabilities), the system cannot appropriately judge the human's input and its actions to prevent undesired or dangerous outcomes. The aim of this proposal is twofold: first, *to define an effective measure of robotic trust* and second, *to evaluate different methods of applying robotic trust within the domain of human-robot control sharing.*

Related Work: Recent trends in rehabilitative and assistive robotics (~10 years) demonstrate an evolution from developing fully autonomous systems that complete tasks for users towards robots that work alongside users and allow the human partner to retain as much control as possible.[2] Researchers have developed numerous methods for allocating control between the human and the robot including linearly blending human and robot control inputs,[3] dynamically shifting between defined levels-of-autonomy,[4] and data-driven probabilistic methods that predict when to provide assistance.[5,6] Literature has recently shown that dynamic autonomy allocation methods are not always optimal to the individual and thus require further customization.[7] Furthermore, the majority of shared control literature assumes the human partner exhibits a static level of control competency, an assumption that is invalid in many scenarios (e.g. skill improving via experience or skill degrading with injury or fatigue).

If a person is distracted or fatigued, they may provide unreliable control inputs that may result in unintended and possibly dangerous consequences.[1] According to [8] robotic trust is a measure of human cogency; the more reliable a human's control signal, the higher the measure of robotic trust. In this work, robotic trust is measured as the difference between a user's reference control trajectories and optimal control solutions.[9] Further, this formulation of robotic trust is easily portable to other robotic platforms. Other formulations of trust compare human and robot performance using robot-specific measures and allocate or remove complete control authority as a direct result of trust measures.[10] Further, these methods attempt to model human autonomy switching patterns for use in model predictive control. Few methods presently exist that quantify the human's cogency when issuing commands in the human-robot team. Therefore, these articles serve as the primary basis for the development of robotic trust.

Technical Aim 1: Identify effective measures of robotic trust based on human inputs. Three initial definitions of robotic trust will be analyzed: **divergence, comprehension,** and **performance**. Inspired by [9], which defined robotic trust as the difference between optimal control trajectories and a human control inputs, **divergence** will inspect the difference between the human's control inputs and the robot's trajectories. Computations of divergence may be simple (e.g. the dot product) or complex (e.g. the Fréchet distance).[9] **Comprehension** will quantify the human's understanding of their robotic partner's physical limitations via control theoretic system stability metrics[11-12] and safety policy violations (e.g. frequency of autonomous interventions).[13] Finally, **performance** uses task-agnostic means (e.g. efficiency measures such as mean completion time)[14] to quantify the human's ability to effectively accomplish an assigned task. An effective trust measurement may be a combination of all three of the aforementioned measures. Combinations of the methods may aggressively select the best performing trust metric, temporally average trust metrics, or weigh trust metrics in a blending scheme where weights are learned to maximize overall trust for an individual or task.

Evaluation: Robotic trust measures will be explicitly calculated while individuals complete a variety of tasks with different robotic platforms. Both uninjured and differently-abled persons will participate in these studies. Trust measures will be compared to a variety of online and offline performance measures to determine which measures are most effective. Such performance measures may include post task user surveys, online user feedback (e.g. a request for assistance), online researcher feedback (e.g. annotations when users fail to recognize important safety features), or trial completion times.

Technical Aim 2: Applications of robotic trust measures to improve shared control. Recent trends in assistive and rehabilitation robotics seek to maximize the human's control of their robotic partner while simultaneously ensuring safety and efficiency.[2,7] In **linear autonomy blending**,[13] trust may be used to directly allocate the autonomy blending parameter combining human and robot control. In **discrete autonomy allocation**, the system measures the performance of the human partner through pertinent measures in real-time, and can dynamically adjust robotic trust to decide when to shift between discrete autonomy levels[4] (e.g. obstacle avoidance, full autonomy, etc.). In **probabilistic blending,** trust is used to modulate control authority between the human and robotic partners; here, trust is modeled by a probability distribution that is continuously updated to incorporate the most recent information[15] (e.g. the robot state, environment state, and/or task information). A single application of trust may not be sufficient for an entire task. Recent literature indicates the need to break some tasks into subtask primitives[3,16-17] for control sharing. This suggests development of **task-dependent trust** measures and applications. Intent inference models[18] can be used to predict the human's task, break it into subtasks, and dynamically shift between appropriate measures of robotic trust. This aim further seeks to explore and define both task-agnostic and -dependent trust applications.

Evaluation: Users, both healthy and differently abled, will be asked to complete a series of tests both with and without trust-tempered robotic assistance on a variety of robotic platforms. Comparing user performance and preferences, with and without robotic assistance, will inform which applications of robotic trust are most effective. Tests will also be designed to classify task-specific trust applications. Measures of human-robot performance may be temporal (completion time), spatial (accuracy of placement or arrival), or derived from safety (number of experimenter interventions). Useful measures of trust effectiveness may also include cognitive loading or perceived task difficulty. The NASA MATB-II "distraction task"[19] and the NASA Task-Load Index Survey[20] will be used to measure cognitive loading and perceived difficulty, respectively. Prior to robotic platform evaluations, subjects will validate robotic trust measures and applications through robotic simulations built using Gazebo.[21]

Motivational Domain: The fundamental constraints faced by both individuals with physical impairments and warfighters makes assistive and rehabilitative robotics an extremely salient domain for the development of robotic trust measures. In both combat environments and rehabilitative settings, the human-robot team is heterogeneous, the human is limited in their abilities to provide control signals, and the cost-of-error is extremely high. Additionally, the reliability of control inputs from differently abled individuals must be maximized for their continued mobility. While combat environments are more dynamic, those in rehabilitative environments are further constrained by their ability to provide command signals due to impairment. As a disease progresses, rehabilitation improves ability, or an individual's assistance preferences change, autonomy must adapt its level of trust to meet the needs of the individual. This parallels combat situations where the robotic partner must consider human ability when adapting to adversarial threats invisible to humans or to situations where skilled warfighter may wish to rely on their intuition and override a machine in order to properly execute a mission. Finally, this work poses to significantly improve the lives of America's impaired veterans.

Resources: The argallab at Northwestern University is uniquely situated to develop measures and applications of robotic trust. The group has a well-established track record working within assistive and rehabilitative robotics.[1,7-9,13,17] The lab has a custom, autonomous wheelchair platform (Fig. 1), a one-of-a-kind wheelchair obstacle course (Fig. 2) and two robotic arm platforms (Fig. 3) from Kinova Robotics,[22] each of which can be used to validate various trust formulations on real hardware. Furthermore, the argallab is situated within the Shirley Ryan AbilityLab, the United States' premier rehabilitation hospital. This facility permits robotic trust studies with numerous patient populations and collaborations with world-leading clinicians. Moreover, a large majority of recent research in robotic trust has been conducted at Northwestern University by argallab Director Prof. Brenna Argall and collaborator Prof. Todd Murphey, allowing for access to expert mentorship and insight.

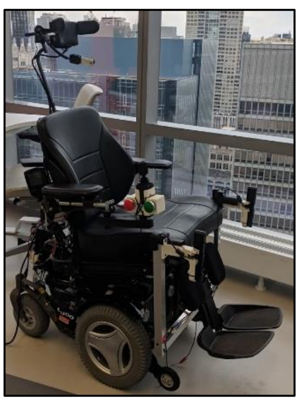

Fig 1: The custom autonomous wheelchair is composed of a Permobil C300 wheelchair, a computer running the Robotic Operating System (ROS),[23] two ASUS Xtion RGB-D cameras,[24] and a variety of control interfaces (joystick, head array, and sip-and-puff device).

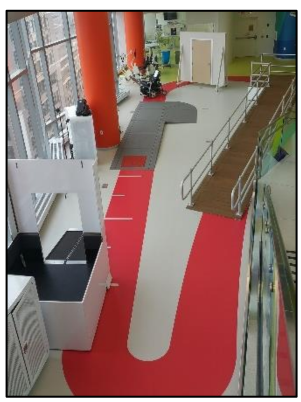

Fig 2: The wheelchair obstacle course was custom designed to simulate common challenges faced by wheelchair users.

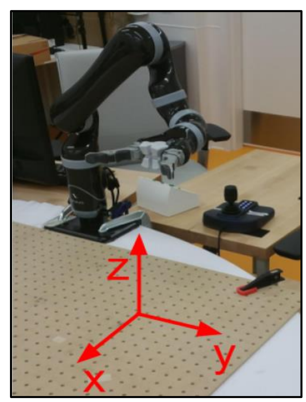

Fig 3: The MICO 6-DOF robotic arm by Kinova Robotics[25] is one of two argallab-owned robotic arms been adopted by users worldwide as both assistive and research platforms.[25] Kinova arms are equipped with onboard sensors and out-of-the-box ROS compatibility.

Works Cited:

1. M. Young, *et al.,* An Analysis of Degraded Communication Channels in Human-Robot Teaming and Implications for Dynamic Autonomy Allocation. In: *Field and Service Robotics. Springer Proceedings in Advanced Robotics*, vol 5. Springer. 2018.
2. O. Horn, "Smart wheelchairs: Past and current trends," *2012 1st International Conference on Systems and Computer Science (ICSCS)*, Lille, 2012, pp. 1-6. 2012.
3. S. Musić, *et al.*, Control sharing in human-robot team interaction, *Annual Reviews in Control*, Volume 44, 2017, Pages 342-354, ISSN 1367-5788.
4. M. Chiou, *et al.*, "Experimental analysis of a variable autonomy framework for controlling a remotely operating mobile robot," in *Proceedings of the IEEE/RSJ International Intelligent Robots and Systems (IROS)*, 2016.
5. A. Kucukyilmaz *et al.*, "Learning Shared Control by Demonstration for Personalized Wheelchair Assistance," in *IEEE Transactions on Haptics*, vol. 11, no. 3, pp. 431-442, 2018.
6. M. Chen, *et al.* 2018. Planning with Trust for Human-Robot Collaboration. In *Proceedings of 2018 ACM/IEEE International Conference on Human-Robot Interaction, (HRI '18).* 2018.
7. D. Gopinath, *et al.*, "Human-in-the-Loop Optimization of Shared Autonomy in Assistive Robotics," in *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 247-254, Jan. 2017.
8. B. D. Argall, *et al.* Computable trust in human instruction. In *Artificial Intelligence for Human-Robot Interaction - Papers from the AAAI Fall Symposium, Technical Report*. 2014.
9. A. Broad, *et al.*, "Trust Adaptation Leads to Lower Control Effort in Shared Control of Crane Automation," *in IEEE Robotics and Automation Letters*, vol. 2, no.1, pp. 239-246, Jan. 2017.
10. H. Saeidi *et al*., "Trust-based mixed-initiative teleoperation of mobile robots," *2016 American Control Conference (ACC)*, 2016.
11. G. F. Franklin, *et al.*, *Feedback control of dynamic systems*, 7th ed. Boston: Pearson, 2015.
12. X. Xu, *et al.*, Robustness of Control Barrier Functions for Safety Critical Control, IFAC-PapersOnLine, Volume 48, Issue 27, 2015, Pages 54-61, ISSN 2405-8963.
13. A. Erdogan, *et al.,* The effect of robotic wheelchair control paradigm and interface on user performance, effort and preference…, in *Robotics and Autonomous Systems*, 2017.
14. A. Steinfeld, *et al.* Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction (HRI '06)*. 2006.
15. S. Reddy, *et al.* Shared Autonomy via Deep Reinforcement Learning. In *Proceedings of Robotics: Science and Systems (RSS '18),* 2018.
16. X. Yang, *et al*, "A framework for efficient teleoperation via online adaptation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017.
17. M. Young, **C. Miller,** *et al.,* "Implications of Task Features in Robotic Manipulation for Dynamic Autonomy Allocation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017. (In Review).
18. S. Javdani, *et al.* Shared autonomy via hindsight optimization for teleoperation and teaming. In *The International Journal of Robotics Research*. 2018.
19. S. Hart, *et al. D*evelopment of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology.* Volume 52. 1988, Pages 139-183.
20. J. Comstock, *et al.* "The multi-attribute task battery for human operator workload and strategic behavior research." Hampton, VA, NASA Langley Research Center. 1992.
21. "Gazebo." gazebosim.org. 2018.
22. "Kinova." www.kinovarobotics.com. 2018.
23. "The Robotic Operating System." ros.org. 2018.
24. "ASUS Xtion PRO." asus.com/3D-Sensor/Xtion_PRO/. 2018.
25. V. Maheu, *et al.* Evaluation of the jaco robotic arm: clinico-economic study for powered wheelchair users with upper-extremity disabilities. In *International Conference on Rehabilitation Robotics (ICORR '11)*, 2011.