Towards a Task-based Metric for Measuring Trust in Autonomous Robots for Everyday Activities

RACHEL RINGE and ROBERT PORZEL, Universität Bremen, Digital Media Lab, Germany

The success and efficiency of collaborative tasks are heavily influenced by trust, especially in household environments, where mistakes by either party could lead to injuries in addition to failure. With the development of artificial agents for complex household tasks moving forward, it becomes increasingly important to explore the amount of trust put in these autonomous agents and its development. To measure trust in everyday household activities, we propose a metric based on the analysis of delegated tasks during a cooperative everyday activity, such as cooking. We consider the aspects of difficulty, involved risk, possibility of error mitigation as well as supervision by the human to be indicators for the level of trust. We also considered discount factors that might convince a human to delegate a task despite a lack of trust in the autonomous agent. In addition, we present a Virtual Reality test environment that can be used to observe humans during cooperative cooking with an artificial partner together with an evaluation approach for the proposed metric using the test environment.

CCS Concepts: • **Human-centered computing** → *Interaction paradigms*; • **Computer systems organization** → *Robotics*.

Additional Key Words and Phrases: human-robot interaction, human-robot trust, trust measure

ACM Reference Format:
Rachel Ringe and Robert Porzel. 2023. Towards a Task-based Metric for Measuring Trust in Autonomous Robots for Everyday Activities. In . ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Trust is an important factor for the success and effectiveness of a collaborative task in which the participants rely on each other to achieve specific sub-goals. For example, in household environments, such as a kitchen, mistakes can be made by either party that could not only lead to failure to complete the task, but even to injury through various hot or sharp appliances. Trust in a new system or technology is critical to its success, since people tend to employ systems that they trust, and reject systems that they do not trust.

In the last few years, artificial agents, such as vacuuming robots, have become more common in household environments, and assistants for more complex tasks as cooking or cleaning are being developed [1]. To ensure that these new systems will be accepted, it is important to explore how much people trust an autonomous system to handle these tasks, how this trust changes during use and what factors lead to an increase or decrease in trust. Toward that goal, it is important to find applicable measures for trust. In this paper we propose measuring a user's trust in an artificial collaborator during cooperative cooking tasks by analysing the tasks delegated to the artificial partner during collaborative execution of a recipe. Therefore, we present a virtual reality (VR) test environment in which experiments measuring these factors can be conducted.

Manuscript submitted to ACM

2 RELATED WORK

Prior research has focused both on trust in automated systems and specifically robots, as well as on measures of trust for the delegation of tasks to robots. Extensive research has been conducted about the nature of trust and how it influences interactions between humans, as well as between humans and robots and other automated systems. One of the challenges when defining trust in Human-Robot-Interaction is the variety of situations and contexts in which robots are used [5]. Trust in a robot during an evacuation in an emergency scenario would need to be defined differently than trust in a robot during a solely analytical task with no possible danger to the human.

Lee and See define trust as ”the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability” [7] while Wagner sees trust as ”a belief, held by the trustor, that the trustee will act in a manner that mitigates the trustor’s risk in a situation in which the trustor has put its outcomes at risk” [17]. Both definitions share the notion that one party is dependant on another party to reach their goal, which puts them in a vulnerable and risky position. A large amount of research in human-automation trust suggests that it is mostly performance-based, while trust among humans is additionally influenced by the other agent’s perceived moral values and the relationship between the agents. Malle et al suggest that both of these aspects are part of a multidimensional definition of trust since robots more and more take on the roles of social partners as well [9]. However, the amount of dimensions that apply are dependant on the situation, e.g., that types of robots that foster a certain kind of trust might be more suitable for more trust-dependent scenarios [5].

Often measures of trust are based on self-reports by the study participants where they rate their trust in the robot’s capabilities. These questionnaires vary between the simple question ”Do you trust this robot?”, highly situation-specific questionnaires and adjusted psychology questionnaires [3]. Schaefer developed a nuanced scale for trust in human-robot-interaction that can be used in various types of situations [15]. Other possible measures of trust are situation-specific measures, based on user behavior such as:

- following the robot’s guidance in an emergency situation [12]
- degree of the user’s compliance with a robot’s requests [14]
- the frequency of interventions into an autonomous system [18]

These measures of trust can be used to evaluate how different factors influence human trust in robots in research contexts, but also to adapt the behavior of systems at runtime to calibrate user trust to the appropriate level [18] [2].

Trust is a major factor towards the effectiveness of collaboration between robots and humans, especially when it comes to delegation of tasks [4]. Task-delegation during collaborative tasks between humans is influenced by the perceived importance and urgency of the task, as well as the confidence in the abilities of the partner. Novelty of the task, as well as confidence in the ability to mitigate any errors that might arise, play a role in the decision to delegate a task [6]. The term *human supervisory control* describes a relationship between human and automated system, that mirrors the relationship between human supervisors and their human employees [16]. Task-Delegation interfaces implement a type of supervisory control, and allow humans to delegate tasks of their choosing to automated systems which can then determine the best way to fulfill the given task, and then provide the user with feedback after providing it [11].

The transferability of results, knowledge or skills from a Virtual Reality (VR) environment to the real world and the suitability of VR studies as an alternative to in situ or lab studies has been explored in prior research. Mäkela et al found that humans exhibit similar behavior around public displays in a VR environment and the real world equivalent [8]. Paneva at al conducted a study that showed a VR-simulated prototype of a levitation interface yielding similar

	#	y/n
Difficulty		
Amount of different tools/appliances needed for task		▨
Time spent on task in relation to all tasks		▨
Number of terms and axioms needed in ontology		▨
Risk		
Is the task dangerous to the human?	▨	
Is the task dangerous to the robot?	▨	
Possibility of error mitigation		
Is the task irrevocable/non-repeatable?	▨	
Are the ingredients valuable?	▨	
Supervision		
Can the task be parallelized and was it parallelized?	▨	
No supervision by the human	▨	
No intervention by the human	▨	
Discount factors		
Is the task boring?	▨	
Is the task tedious?	▨	
Is the task unpleasant/gross?	▨	
Is the user unable to do the task on their own?	▨	
User-attitude towards the activity		▨

Table 1. Criteria for delegated task that could indicate trust

levels of user experience and engagement as the real prototype [10]. Saffo et al were able to replicate the results of studies on tabletop collaboration and Fitt's law in a VR-environment [13].

Building on this prior work we seek to further research in measuring trust by proposing specific metrics appropriate for the household domain. These proposed metrics will, in turn, be based on the individual task-specific components that are involved in the delegation of the respective tasks.

3 TASK DELEGATION AS INDICATOR FOR TRUST

As the delegation of tasks among humans relies on trust, we propose that the tasks given to the artificial collaborator, e.g. while preparing a meal, can supply information on the level of trust the human has in them. If the human assigns intricate, dangerous or important tasks to the artificial agent, e.g. heating or cutting an ingredient, this would indicate, that they trust this partner to complete the task successfully. Should they only delegate minor tasks to the robot – for example, wiping the counter – it indicates, that the robotic partner is only trusted to fulfill simple tasks where errors could easily mitigated.

Toward the goal of measuring trust based on task delegation, three different aspects of a task that could influence a human's tendency to delegate it were chosen in our approach: difficulty, risk and possibility for error mitigation. In addition, it was deemed relevant if a human would supervise the artificial collaborator during a task or not. These aspects were combined into a basis for a scale, that can be used to determine the level of trust the human put into the artificial partner when delegating this specific task to them. The collection of components for the chosen aspects can be seen in Table 1 and will be explained in the following section.

Difficulty of the delegated task

Since humans tend to delegate challenging tasks when they have trust in the capabilities of a partner, we assume that the difficulty of a task will be a factor in the amount of trust required for it to be delegated. The first chosen component for difficulty of a task was the amount of different tools or appliances needed for the execution of the task that the artificial partner is meant to complete. This is based on the assumption that tasks that use fewer different tools would be less difficult than ones that use multiple tools/appliances. For example the simple task of wiping the counter-top after use would only require a cloth, while frying mushrooms would require using the stove and a pan as well as a spatula. A second measure for difficulty is the time spent on a task in comparison to all tasks completed during the overall activity. A task that uses only few appliances but takes up half the time collectively spent on a given cooking task is likely more complex a task that only takes 20 seconds. An example for this would be that mixing a dough for 3 minutes would be more complex than adding oil to the pan. The last measure for difficulty is the cognitive effort, denoted by the number of concepts and axioms needed for understanding this task in the used ontology. The higher this number is, the more concepts an agent would need to know and understand to complete the task successfully, hence making it more difficult.

Risk Involved in the Delegated Task

Since everyday activities involve various tools and appliances that might pose a danger to any agent, it is likely that tasks involving these objects would only be delegated to the robot if the human trusts it to not harm humans or itself. A task would be considered dangerous to the human if it involves tools, appliances or ingredients that could cut or burn the human, for example knives or hot water. Tasks that would be considered dangerous to the robot involve liquids that could short-circuit its electronics as well as sharp objects that could damage its components.

Possibility for Error Mitigation

During completion of a household activity there are some tasks where errors are easily mitigated and some where an error would prevent activity from being completed successfully. Delegating task that can not be repeated, e.g., by taking additional ingredients and doing it over or reversed by undoing, would indicate, that the human trusts the robot to complete the task successfully, as it is critical to the completion of the overall activity. In addition, the value of the used ingredients or detergents is included, since a task might be repeatable using additional ingredients, but would involve additional costs. An example would be the use of expensive spices like saffron or high-grade cleaning fluids.

Supervision during Task Completion

The aspect of supervision during task completion is also pertinent when considering trust, since constant supervision or even intervention likely implies that the human does not trust the robot to complete the task successfully in an autonomous manner. In addition, it is important to consider if the task could be parallelized. If this is a possibility, but the human does not take advantage of this opportunity to speed up the completion of the activity or to increase the efficiency of the process, it could imply, that the human is trying to be ready in case an intervention is needed and thus does not trust the robot to complete the task correctly.

Discount Factors

In addition to the aspects of a task that would imply trust by the human, there were also discount factors included in the criteria, that might convince a human to delegate a task to the robot despite not trusting him to complete it successfully or correctly. A boring or tedious task might be delegated, since the user would rather take the risk of ordering the robot to fulfill it, instead of having to complete it themselves. Examples for this might be peeling a large amount of potatoes or stirring a pot constantly for 30 minutes. Similarly the benefits of delegating a task that the human would consider uncomfortable or disgusting could outweigh the risks of entrusting it to the robot. Another important aspect are the abilities of the human. If the human is incapable to complete a task themselves, they would likely order the robot to do it for them despite not trusting them, since they have no other choice. A possible example for this would be fetching an object from a high shelf, that the human is unable to reach. In addition, the human's attitude towards cooking would be relevant, since it could influence the perceived importance and complexity of tasks greatly. For example someone inexperienced with cooking might consider sauteing onions an easy task and delegate it to the robot without hesitation, while a more experienced cook would consider it far more difficult.

4 TEST ENVIRONMENT

To observe humans during cooperative cooking with an artificial partner, a VR application was developed in the Unity game engine for use with an Oculus Quest HMD. In this application the user is placed in a kitchen environment together with a virtual robot. The user can interact with various objects in the kitchen by grabbing them with either their hands or the controllers and then complete various cooking tasks by moving them in appropriate ways – e.g., moving a whisk in circular motions through a bowl containing the different ingredients to be mixed. In addition, the user can order the robot to fulfill any of the needed cooking tasks for recipe completion – e.g., portioning a certain amount of an ingredient into a bowl – or some supporting such as cleaning, tidying or fetching objects for the user. For these orders a delegation-type interface is used, where the user orders the robot to fulfill a task in a declarative manner, but is not required to give details on how the task should be completed.

User actions and orders are handled by a highly-accurate physics simulation running in the background of the VR application using the Bullet physics engine, that simulates the correct behavior for objects moved by the user. This involves, for example, that the robot moves autonomously towards objects needed for completion of the order, grasping them and using them in the appropriate ways to complete the task. The robot used in the physics simulation is abstract and only consists of two grippers and a head. This makes it possible to represent the robot in the VR application with any model that possesses two grippers or similar appendages to interact with objects. An example of this is depicted in Figure 1 .

During task completion the virtual robot will only make minor mistakes that naturally occur during certain tasks due to the accuracy of the physics simulation and do not influence the result of the action. An example would be spilling a minor amount of an ingredient on the counter top when portioning it into a container. To ensure that the human does not delegate all tasks to the robot or complete all tasks themselves, the task given to the user will include completing the recipe in as little time as possible, thus necessitating some delegation and collaboration.

Evaluation Approach

It is planned to first determine the weights for the different criteria and discount factors described in Section 3 with a user study in the virtual reality test environment, to operationalize the criteria into a scale that rates the trust in the robot displayed by delegating the individual tasks. This data can then be used to implement a graphical model, e.g. a Bayesian network, which could then be used to predict if a human would trust an AI with a certain cooking task. The correctness of these predictions could, subsequently, be tested and updated with further user studies.

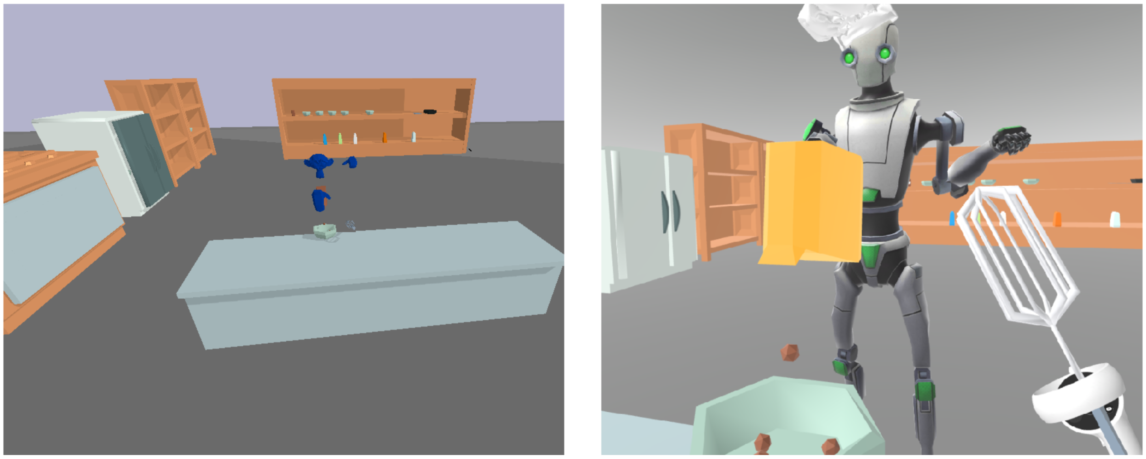

Fig. 1. Comparison of the physics simulation and the VR application showing the artificial robot portioning an ingredient into a bowl while the user is holding a whisk

5 CONCLUSION AND FUTURE WORK

In this work we proposed a scale to rate the trust a human put into an artificial partner during a cooperative household activity. The scale incorporates the difficulty of the task, the risk involved for human and robot, the possibility for error mitigation as well as supervision during task completion. In addition, some discount factors are included, that could influence the human's decision to delegate a task despite insufficient trust in the robot's capabilities. The test environment to conduct experiments using this scale was described as well as an approach to evaluate the results of future experiments.

In the future this scale could be part of a bigger set of measures for trust specific to cooperative tasks, that includes other aspects such as the phrasing of orders given to the robot. Similarly it could be modified for further household tasks, that could in the future be assigned to household robots. Predictions made by a graphical model based on these metrics could also be used to adjust robot behavior at runtime to calibrate trust to the appropriate level for optimal cooperation. The described test environment and scale could be used in the future to explore different robot appearances and behaviors and how they affect trust, as well as trust development over time when the human can observe the robot complete tasks successfully or make errors.

ACKNOWLEDGEMENTS

This work was funded by the by the FET-Open Project #951846 ”MUHAI – Meaning and Understanding for Human-centric AI” by the EU Pathfinder and Horizon 2020 Program and the KI-SIGS Project “AP 390” of the German Federal Ministry for Economic Affairs and Climate Action (BMWK).

REFERENCES

[1] Michael Beetz, Lorenz Mösenlechner, and Moritz Tenorth. 2010. CRAM—A Cognitive Robot Abstract Machine for everyday manipulation in human environments. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1012–1017.

[2] Min Chen, Stefanos Nikolaidis, Harold Soh, David Hsu, and Siddhartha Srinivasa. 2018. Planning with trust for human-robot collaboration. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*. 307–315.

[3] Meia Chita-Tegmark, Theresa Law, Nicholas Rabb, and Matthias Scheutz. 2021. Can you trust your trust measure?. In *Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction*. 92–100.

[4] Amos Freedy, Ewart DeVisser, Gershon Weltman, and Nicole Coeyman. 2007. Measurement of trust in human-robot collaboration. In *2007 International Symposium on Collaborative Technologies and Systems*. 106–114. https://doi.org/10.1109/CTS.2007.4621745

[5] Zahra Rezaei Khavas. 2021. A Review on Trust in Human-Robot Interaction. https://doi.org/10.48550/ARXIV.2105.10045

[6] Katherine J Klein, Jonathan C Ziegert, Andrew P Knight, and Yan Xiao. 2006. Dynamic delegation: Shared, hierarchical, and deindividualized leadership in extreme action teams. *Administrative science quarterly* 51, 4 (2006), 590–621.

[7] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[8] Ville Mäkelä, Rivu Radiah, Saleh Alsherif, Mohamed Khamis, Chong Xiao, Lisa Borchert, Albrecht Schmidt, and Florian Alt. 2020. Virtual Field Studies: Conducting Studies on Public Displays in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376796

[9] Bertram F Malle and Daniel Ullman. 2021. A multidimensional conception and measure of human-robot trust. In *Trust in human-robot interaction*. Elsevier, 3–25.

[10] Viktorija Paneva, Myroslav Bachynskyi, and Jörg Müller. 2020. Levitation Simulator: Prototyping Ultrasonic Levitation Interfaces in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376409

[11] R. Parasuraman, S. Galster, P. Squire, H. Furukawa, and C. Miller. 2005. A flexible delegation-type interface enhances system performance in human supervision of multiple robots: empirical studies with RoboFlag. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 35, 4 (2005), 481–493. https://doi.org/10.1109/TSMCA.2005.850598

[12] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M. Howard, and Alan R. Wagner. 2016. Overtrust of robots in emergency evacuation scenarios. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 101–108. https://doi.org/10.1109/HRI.2016.7451740

[13] David Saffo, Sara Di Bartolomeo, Caglar Yildirim, and Cody Dunne. 2021. Remote and collaborative virtual reality experiments via social vr platforms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[14] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 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*. 141–148.

[15] Kristin E Schaefer. 2016. Measuring trust in human robot interactions: Development of the "trust perception scale-HRI". In *Robust intelligence and trust in autonomous systems*. Springer, 191–218.

[16] Thomas B Sheridan. 2012. Human supervisory control. *Handbook of human factors and ergonomics* (2012), 990–1015.

[17] Alan R Wagner and Ronald C Arkin. 2011. Recognizing situations that demand trust. In *2011 RO-MAN*. IEEE, 7–14.

[18] Anqi Xu and Gregory Dudek. 2015. Optimo: Online probabilistic trust inference model for asymmetric human-robot collaborations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. 221–228.

Received 23 February 2023