Interdependence and Trust Analysis: An Extension to Coactive Design

Carolina Centeio Jorge
Delft University of Technology
Delft, The Netherlands
C.Jorge@tudelft.nl

Catholijn M. Jonker
Delft University of Technology
Leiden University
Delft & Leiden, The Netherlands
C.M.Jonker@tudelft.nl

Myrthe L. Tielman
Delft University of Technology
Delft, The Netherlands
M.L.Tielman@tudelft.nl

ABSTRACT

As machines' autonomy increases, the possibilities of collaboration between human and machine also increase. In particular, different types of interdependence may become feasible. The feasibility of each type of interdependence depends on a complex combination of different dimensions that compose contextual trustworthiness, such as team members' competence (i.e., skills, knowledge), willingness (i.e., intention, preference) and external factors (i.e., opportunity, resources). In this paper, we present an extension of the Coactive Design's Interdependence Analysis table which includes these three dimensions: the Interdependence and Trust Analysis table for human-machine teams. Our table can support the decision-making not only of the team designer, but also of the team members, either human or machine. By using this table as a shared mental model, decisions may become more transparent, justifiable and interpretable, which may lead to an increased and appropriate trust among teammates. We present the work in progress in the evaluation of this table through different focus groups. Preliminary results show that our table is needed by researchers in human-machine collaboration fields.

CCS CONCEPTS

• **Human-centered computing → Interaction design process and methods**.

KEYWORDS

Interdependence, Human-Machine Teamwork, Trust, Trustworthiness, Human-Machine Collaboration

ACM Reference Format:

Carolina Centeio Jorge, Catholijn M. Jonker, and Myrthe L. Tielman. 2023. Interdependence and Trust Analysis: An Extension to Coactive Design. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In scenarios where humans and machines collaborate, several decisions have to be made, such as who does what [1, 2]. In some situations this may be straightforward, such as when there is no overlap of teammates' (human's or machine's) expertises, e.g., imagine a kitchen robot that only works as pressure cooker and a human (who, of course, cannot work as pressure cooker) who can prepare the ingredients that go in the machine. On the other hand, there are situations where both teammates can do certain tasks, for example a kitchen robot arm that also chops vegetables and a person who can do the same. As the autonomy of these machines increases, situations as the latter become more frequent. Along with more capabilities and autonomy, there may come more options of how to design human-machine collaborations, with different types of interdependence involved.

In [9], Johnson et al. present the Interdependence Analysis table as a tool for Coactive Design, i.e., an approach to address the increasingly sophisticated roles that people and machines play as the use of machines expands into new, complex domains. The original Interdependence Analysis lists down, through a table, the different capacities required for the execution of tasks and allows a comprehensive analysis of which teammate has or not such capacities. After filling in such table, one should be able to understand the necessary interdependencies for each task, through a colour code and requirement gathering.

Although a thorough analysis of capacities is an important step, we claim that it is also important to consider other dimension that lead to the success of a task. For a cognitive agent, either human or artificial, to successfully perform a task, they need to have the capacities/capabilites (i.e., "can they do it?"), the willingness/intention to do it (i.e., "will they do it"?) and to have the external opportunities/permissions to do it (i.e., "is it possible to do it"?) [5, 6]. These dimensions, in other words, compose the trust that an agent has on its teammate to successfully perform a task. In [8], Johnson et al. already propose an extension of the table that includes trust as one extra dimension to consider when analysing the interdependences. However, this dimension of trust (1) is only considered for the trust in one of the agents (the performer) involved in the interdependence, and, in our opinion, (2) could be further divided into dimensions that are easier to assess and update.

In this paper, we present our version of the Interdependence Analysis table, the *Interdependence and Trust Analysis table* and the preliminary results of its evaluation. This table can be used by the team members performing the task, either human or artificial, or by a team designer. Furthermore, it can be used as a decision-making support system, as well as a shared mental model [13, 14, 16], which increases transparency among teammates and facilitates justification of one's actions, facilitating appropriate trust [15, 17].

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 INTERDEPENDENCE AND TRUST ANALYSIS

The goal of our proposed analysis is twofold. Firstly, we want a tool that provides a more comprehensive analysis of all possible team configurations based on the feasibility of the interdependences at the atomic task (i.e., a task that is not composed of sub-tasks) level. Moreover, for each of these interdependences, we want a tool that analyses the trustworthiness of each teammate for a certain role. For this, we analyse not only the competence/performance dimension, but also the willingness/intention, as well as the external factors that may restrain that action.

2.1 Trust

Our tool can help to analyse a team member's trustworthiness for a role in a certain task, within a certain context. With such tool, either a team designer or other team members can more easily choose when to rely on whom. Trustworthiness is a complex concept, and following the literature it can consist of a set of dimensions that range from the trustee's competence to its intentions [7]. Models in slightly different settings propose that trust depends on how one perceives another's 1) Ability, Benevolence and Integrity [12] (in human organizations), 2) Willingness, Competence and Dependence [3] (in multi-agent systems), and 3) Performance, Process and Purpose [10] (when the human is the trustor and an artificial agent is the trustee). The way trustworthiness is perceived can also depend on trustor's characteristics [12] and is usually influenced by external factors, which are contextual conditions determining the situation in which the task is executed [6], such as environmental configuration, emotional state, workload, etc. Based on the literature and studies we have run before in human-machine teams (in publication process), we include (1) a belief related to ability, performance, competence (column "can" in Figure 2), (2) a belief which comprehends everything besides ability that may contribute to the choice of performing a task successfully, i.e., willingness, benevolence, integrity, and (3) the context which comprehends external factors (opportunities, permissions), in our analysis.

Team configuration	Joint	H	M
human performer + no support		✓	
human performer + machine support	✓	✓	
mandatory joint	✓		
machine performer + human support	✓		✓
machine performer + no support			✓

Figure 1: Team configurations based on interdependence and trust analysis in Fig. 2. Each configuration requires at least those positive assessments. The unfilled grey cells can be either positive (✓) or negative (X).

Possible Performer	Can	Will	External Factors	F
Joint	✓	✓	✓	✓
H independent	X	✓	✓	X
M independent	✓	✓	X	X

Figure 2: Interdependence and Trust Analysis Table for one atomic task. Possible performers are the human (H), the machine (M) or both being co-performers (Joint). To each of these possibilities, we analyse whether they have the skills and knowledge to do a task (whether that performer *can*), whether they have the intention and preference to do the task (the performer *will*) and, finally, if the external factors and permissions allow. We can see the resulting feasibility (F) of each performer option in the last column.

2.2 Interdependence

For each possible interdependence, we need to estimate the trustworthiness of the teammates for the particular role they would take in the interdependence. We consider five different types of team configuration. If we consider independent configurations, we can have either a completely independent human performer, or a completely independent machine performer. There are also two possible soft interdependencies, i.e., human with support (this happens when the human can be independent, but support is possible to increase efficiency or reliability), and machine performer with human support and machine as an independent performer. Finally, there is also a hard interdependence, i.e., mandatory joint (human and machine have to perform the task together).

Ideally, we would assess competence (can), willingness/intention (will) and external factors for each teammate (2 x 3), for each possible interdependence configuration (x 5). However, this proved to be quite overwhelming for human participants (presented in Section 3, so instead we analyse only the possible performers instead, and infer the support feasibility. In particular, for each task, we consider having a joint performer (human and machine as co-performers), or the human as performer (H independent), or the machine as performer (M independent). We infer the supporting roles given that joint is possible, i.e., if joint is possible, support is also possible, as in Figure 1.

2.3 The table

Figure 2 presents the focal part of the interdependence and trust analysis for an atomic task to be performed by a team composed of a human (H) and a (partially) autonomous machine (M). For each possible performer, the user can fill in the table with "✓" if positive or with "X" if negative. For example, when we analyse whether the human can perform the task independently, we should consider whether they can (i.e., have the competences, skills, knowledge...), whether they will (i.e., want to, would choose to do that task) and, finally, if they have the external opportunities and resources to do it (i.e., external factors).

Let's imagine that we have the subtask of analysing some type of medical scans, and we want to analyse the feasible interdependences. We start by filling in the table with the constraints we know of. In this case, we can imagine that the human can't analyse alone, as it is know that it does not have the required the capacities, but they can be still helpful to increase reliability or efficiency. We also know that the machine is not allowed to analyse alone, as it is required by legislation that a human also has a look. This leads to the crosses ("X") in Figure 2. It seems that the only feasible performer is the joint one, which leads to "mandatory joint" being the only feasible configuration (as per Figure 1). Once the table in Figure 2 is completed for all sub-tasks, i.e., there are other tasks such as writing the report of the scans in natural language, etc, the user can choose which team configuration is more adequate for each atomic task, based on the information in Figure 1.

2.4 Use of the Table

This table can not only be used by a team designer, with an overview of all tasks and teammates, but also by the teammates themselves, either human or artificial. In particular, it can be used by the machine to make decisions on whether to support the human or not, for example, or who to call for help for a certain task (see more in [4]). If the machine can update its beliefs of other teammates and itself according to this structure and representation, this tool can be used to ensure transparency and potentiate justifications from the artificial teammate. This tool also offers a good analysis of dyadic (and possibly team) trust, which can facilitate appropriate and warranted trust among teammates [11].

3 EVALUATION

The table has been iterated several times, in light of the experts' feedback gathered on focus groups, and needs to be further evaluated. More concretely, we divided the evaluation of our method in two parts. The first part was intended to improve the table, with two small expert groups, each composed of two researchers in the field of human-machine interaction and collaboration. With the completion of the first part, we reached to the version of the table presented in this paper. The second part is aimed at evaluating its final usability with a use case in the domain of firefighting. By the time we are writing this paper, this study is planned to happen soon.

Each focus group lasted one hour and a half and was composed of (1) analysing interdependence and trust of a collaborative task (to be executed by a team composed of one human and one machine) by filling in the new version of the table, and (2) answering six open questions. The collaborative task was inspired by the video game Overcooked and consisted of preparing and frying vegetables. The subtasks were "washing the veggies", "chopping the veggies", and "frying the veggies". The participants were informed of several constraints to the successful execution of these tasks.

Participants were told that the machine could not hold a knife, which, in the current version of the table (different from the version presented in these focus groups), should make the cells of "external factors" negative ("X") both for *M (machine) independent* and *Joint*. They were also told that the human did not want to fry veggies alone, or wash the veggies jointly. This information should make "Will" negative ("X") for *H (human) independent* in "frying the veggies" and for *Joint* in "washing the veggies". Finally, we also shared constraints regarding competences, for example, the human teammate did not know how to fry, which should lead to a negative ("X") "Can" for this subtask in both *H (human) independent* and *Joint*.

The questions we asked participants at the end included "What is the one thing you liked best/least?", "What would you change/keep in the table?" and "In which situations would you use or not use the table?". Most participants showed great interest in using our table for their personal research works. Among other things, participants mentioned our table would be useful in the process of designing their experiments' tasks, calibrating appropriate trust between humans and machines, and designing explanations. We have received negative feedback mainly based on the colour code, the efficiency related to filling in the table, and possible overlapping of dimensions. Furthermore, it was clear that filling in the first iteration of the table was quite overwhelming for human participants. All this feedback was already integrated in the version of the table presented in this paper. This study was previously approved by the University's Human Research Ethics Committee (HREC).

4 CONCLUSION

In this paper, we present an extension of the Interdependence Analysis for human-machine teams. Our approach includes a discriminated analysis of the trustworthiness dimensions of competence (i.e., skills, knowledge), willingness (i.e., intention, preference) and external factors (i.e., opportunity, resources), for each possible team interdependence configuration, for each subtask. This table can support the decision-making not only of the team designer, but also of the team members, either human or machine. By using this table as a shared mental model, decisions may become more transparent, justifiable and interpretable, which may lead to an increased and appropriate trust among teammates. In future work, we want to implement a decision-making model on the machine teammate that makes use of the table structure and updates it throughout interactions.

ACKNOWLEDGMENTS

This work is part of the Delft AI Initiative - AI*MAN lab. Thanks to all focus group participants and Matt Johnson for the feedback.

REFERENCES

[1] Arsha Ali, Hebert Azevedo-Sa, Dawn M Tilbury, and Lionel P Robert Jr. 2022. Heterogeneous human–robot task allocation based on artificial trust. *Scientific Reports* 12, 1 (2022), 15304.
[2] Hebert Azevedo-Sa, X Jessie Yang, Lionel P Robert, and Dawn M Tilbury. 2021. A unified bi-directional model for natural and artificial trust in human–robot collaboration. *IEEE robotics and automation letters* 6, 3 (2021), 5913–5920.
[3] Cristiano Castelfranchi and Rino Falcone. 2010. *Trust & Self-Organising Socio-technical Systems*. Springer International Publishing. 209–229 pages. https://doi.org/10.1007/978-3-319-29201-4_8
[4] C Centeio Jorge, CM Jonker, and ML Tielman. 2023. Artificial Trust for Decision-Making in Human-AI Teamwork: Steps and Challenges. In *CEUR Workshop Proceedings*, Vol. 3456.
[5] Rino Falcone and Cristiano Castelfranchi. 2004. Trust Dynamics: How Trust Is Influenced by Direct Experiences and by Trust Itself. In *3rd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004), 19-23 August 2004, New York, NY, USA*. IEEE Computer Society, 740–747.
[6] Rino Falcone, Michele Piunti, Matteo Venanzi, and Cristiano Castelfranchi. 2013. From manifesta to krypta: The relevance of categories for trusting others. *ACM Trans. Intell. Syst. Technol.* 4, 2 (2013), 27:1–27:24. https://doi.org/10.1145/2438653.2438662
[7] N. Griffiths. 2005. Task delegation using experience-based multi-dimensional trust. In *AAMAS '05*.

[8] Matthew Johnson and Jeffrey M Bradshaw. 2021. The role of interdependence in trust. In *Trust in human-robot interaction*. Elsevier, 379–403.

[9] Matthew Johnson, Jeffrey M Bradshaw, Paul J Feltovich, Catholijn M Jonker, M Birna Van Riemsdijk, and Maarten Sierhuis. 2014. Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction* 3, 1 (2014), 43–69.

[10] J. D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of Human Factors and Ergonomics Society* 46 (2004), 50 – 80.

[11] Michael Lewis, Huao Li, and Katia Sycara. 2020. Deep learning, transparency, and trust in human robot teamwork. In *Trust in Human-Robot Interaction*. Elsevier, 321–352.

[12] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An Integrative Model of Organizational Trust. *Source: The Academy of Management Review* 20 (1995), 709–734. Issue 3.

[13] Eduardo Salas, Dana E Sims, and C Shawn Burke. 2005. Is there a "big five" in teamwork? *Small group research* 36, 5 (2005), 555–599.

[14] Matthias Scheutz, Scott A DeLoach, and Julie A Adams. 2017. A framework for developing and using shared mental models in human-agent teams. *Journal of Cognitive Engineering and Decision Making* 11, 3 (2017), 203–224.

[15] Anna-Sophie Ulfert, Eleni Georganta, Carolina Centeio Jorge, Siddharth Mehrotra, and Myrthe Tielman. 2023. Shaping a multidisciplinary understanding of team trust in human-AI teams: a theoretical framework. *European Journal of Work and Organizational Psychology* (2023), 1–14.

[16] Iris van de Kieft, Catholijn M Jonker, and M Birna van Riemsdijk. 2011. Shared mental models for decision support systems and their users. In *3rd International Workshop on Collaborative Agents-REsearch and development (CARE 2011)*. 54–63.

[17] Ruben S Verhagen, Siddharth Mehrotra, Mark A Neerincx, Catholijn M Jonker, and Myrthe L Tielman. 2022. Exploring Effectiveness of Explanations for Appropriate Trust: Lessons from Cognitive Psychology. *arXiv preprint arXiv:2210.03737* (2022).