Proceedings Paper:
Cameron, D. orcid.org/0000-0001-8923-5591 and Collins, E.C. (2021) User, robot, deployer : a new model for measuring trust in HRI. In: 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) workshop on Trust, Acceptance and Social Cues in Human-Robot Interaction (SCRITA 2021). SCRITA 2021 : Trust, Acceptance and Social Cues in Human-Robot Interaction, 12 Aug 2021, Virtual conference. arXiv.org .

Physical, Psychological, Team, Organisational, and Political levels), which all connect. Given this, we propose that whilst the lower levels of Physical, Psychological and potentially even Team are relatively well-studied in understanding trust in HRI, the under-considered higher levels of Organisational and Political also impact trust in even small-scale interactions.

II. THE SOCIAL TRIAD OF HRI

While the commonly-held approach to HRI as interaction between user and robot contained within a specified scenario may reliably capture key elements of interest, this view marginalises the influence higher levels, crucially the person/agent deploying the robot to the user, have on the 'shape' of the scenario and the interaction experience. Where a programmer, owner, or researcher using a robot for HRI may consider themselves to be external to the scenario, their influence may still be experienced by the user (the 'H' in HRI). Our model (Fig. 1) seeks to explicitly include the role that these 'external' agents may have on HRI, and their impact on measurements of trust in the interaction.

Building on the approach to trust in HRI from a social perspective, we expand the common view of HRI as a social dyad between User and Robot to view HRI as a social triad that includes the Deployer. 'Robot' and 'User' as terms should be familiar to a researcher in HRI; for this model, 'Deployer' is the interested external agent, such as a researcher or manager, and/or body, such as a corporation, who brought the scenario into being. The User's perceived relationship with the Deployer - both any literal relationship and their regard for the deployer as representing the Organisational and Political decisions in creating the scenario - may influence their views towards the scenario and interactions with the robot.

A. Interactions in the Triad

Perhaps the most apparent interaction for study in the social triad is the (often reciprocal) one between the User and the Robot (i.e. HRI). A user may influence a robot's actions through their own behaviour, either directly controlling or otherwise affecting the robot. In return, users' experiences and their evaluations of the interaction may be influenced by social-like behaviours of a robot. Specifically, the trust an

User, Robot, Deployer: A New Model for Measuring Trust in HRI

1st David Cameron
Information School
The University of Sheffield
Sheffield, UK.
d.s.cameron@sheffield.ac.uk
ORCID 0000-0001-8923-5591

2nd Emily C. Collins
Department of Computer Science
University of Manchester
Manchester, UK.
e.c.collins@manchester.ac.uk
ORCID 0000-0001-9396-536X

Abstract—There is an increasing interest in considering, implementing, and measuring trust in human-robot interaction (HRI). Typically, this centres on influencing user trust within the framing of HRI as a dyadic interaction between robot and user. We propose this misses a key complexity: a robot's trustworthiness may also be contingent on the user's relationship with, and opinion of, the individual or organisation deploying the robot. Our new HRI triad model (User, Robot, Deployer), offers novel predictions for considering and measuring trust more completely.

Index Terms—human-robot interaction, trust, transparency, deployer, user

I. INTRODUCTION

Research aimed at measuring trust in Human-Robot Interaction (HRI) scenarios is fundamental to the development of readily adoptable robotics and autonomous systems (RAS) technologies. Both social and industrial RAS need to be presented to users as approachable, efficient, and safe. Existing research commonly explores trustworthy HRI via a dyadic approach: labelling an individual user as one half of an interaction, and attempting to optimise a trust benchmark outcome from said individual both during and after an interaction with a robotic agent [1]–[3]. Indeed, there are a range of identified factors - from the robot itself, the user, and the environment - that may affect trust measurement in HRI [2]. For example, the use of social/social-like behaviour expressed by a robot is considered as a factor against which a user's trust in that particular robot is measured [4]. However, in this position paper we argue that this range of factors is incomplete. Borrowing from Vicente's model of the Tech Ladder [5], we propose that measuring trust in HRI requires a triadic approach, in which the deployer of the robot, and the relationship said deployer has with the robot-user, is also considered as an impacting factor.

A. HRI across multiple levels

Vicente writes that technology is *any* human invention to meet a human or societal need [5]. Where RAS and Power-plants are examples of hard (i.e. tangible) technology, other inventions such as Staff-Workload-Allocation and Democracy are conversely soft technologies. Interactions with deployed robotics can be studied at varying levels of abstraction (the Physical, Psychological, Team, Organisational, and Political levels), which all connect. Given this, we propose that whilst the lower levels of Physical, Psychological and potentially even Team are relatively well-studied in understanding trust in HRI, the under-considered higher levels of Organisational and Political also impact trust in even small-scale interactions.

II. THE SOCIAL TRIAD OF HRI

While the commonly-held approach to HRI as interaction between user and robot contained within a specified scenario may reliably capture key elements of interest, this view marginalises the influence higher levels, crucially the person/agent deploying the robot to the user, have on the 'shape' of the scenario and the interaction experience. Where a programmer, owner, or researcher using a robot for HRI may consider themselves to be external to the scenario, their influence may still be experienced by the user (the 'H' in HRI). Our model (Fig. 1) seeks to explicitly include the role that these 'external' agents may have on HRI, and their impact on measurements of trust in the interaction.

Building on the approach to trust in HRI from a social perspective, we expand the common view of HRI as a social dyad between User and Robot to view HRI as a social triad that includes the Deployer. 'Robot' and 'User' as terms should be familiar to a researcher in HRI; for this model, 'Deployer' is the interested external agent, such as a researcher or manager, and/or body, such as a corporation, who brought the scenario into being. The User's perceived relationship with the Deployer - both any literal relationship and their regard for the deployer as representing the Organisational and Political decisions in creating the scenario - may influence their views towards the scenario and interactions with the robot.

A. Interactions in the Triad

Perhaps the most apparent interaction for study in the social triad is the (often reciprocal) one between the User and the Robot (i.e. HRI). A user may influence a robot's actions through their own behaviour, either directly controlling or otherwise affecting the robot. In return, users' experiences and their evaluations of the interaction may be influenced by social-like behaviours of a robot. Specifically, the trust an
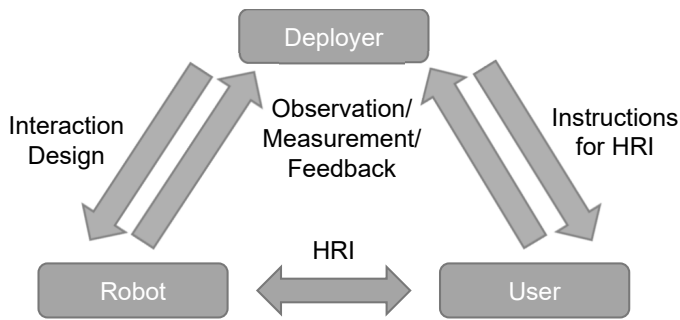
Fig. 1. Pathways of communication for an HRI scenario

individual has towards a robot may be affected by the robot's appearance and behaviours [2] and social strategies to regulate trust [6]–[8].

The interaction from Deployer (D) to Robot (R) includes the production of robot behaviours through direct control of a robot (e.g., scenarios requiring Wizard of Oz control [9]), specifying goals for the robot, or developing architectures to generate behaviour [10]. While the Deployer may be responsible for programming the robot, it is not necessary for them to be present to have influence in the social triad. The Deployer specifies the contexts for which robots may be used and the bounds for a robot's interaction with the User (U) [11].

D to U: Like the Deployer determines the bounds of a Robot's behaviour, their construction of an interaction scenario may constrain the User. Communication from the Deployer and existing or emerging relationships between Deployer and User may shape a User's approach to HRI: Users might evaluate a scenario based on views of the Deployer's behaviours and intentions while ostensibly evaluating the robot.

U to D: feedback on the scenario - either passively when under observation or actively through agreed feedback channels (in research, this may be questionnaires, interviews etc.; in industry, this may be performance appraisals etc.).

R to D: Sharing of information on the interaction (passively through the behaviour) or actively through recorded metrics from the interaction, potentially of both Robot and User.

Thus, where Users may need to trust Robots, they also need to trust Deployers (e.g., Deployers will not allow the User to come to harm (physical or psychological); they have programmed the Robot to operate appropriately (and transparently); they have specified an HRI scenario which complies with necessary safety/ethical regulations; they do not use information gathered from HRI to adversely affect Users.)

B. Interaction Scenarios

What the model offers: HRI is a scenario into which a user has entered but not-necessarily specified. Trust in HRI may depend on the communication and trust towards the deployer as well as trust towards the robot; any measurement of trust the user may have of the robot may also reflect how much trust the user has in the deployer. Further, the model can be used to highlight the importance of transparency in understanding trustworthy HRI. As well as knowing the processes involved in an HRI scenario, a user also wants to know if their interaction is going to be reported (e.g., is the robot serving as a secondary channel for recording the user beyond any direct interaction from U to D?). Although a user can withdraw from an HRI scenario, given the role of D, they may have relatively little power to shape or define the HRI scenario itself. Thus, crucially, the user, at the whim of the deployer, may provide a trust assessment of the robot that is biased due to the influence of the deployer.

III. PROPOSED DIRECTIONS FOR FUTURE STUDY

In sum, our model predicts:

1) The perceived trustworthiness of the deployer is positively associated with trust in interactions.
2) Participant's articulation of trustworthiness that includes views towards the deployer would explain some variation in user-differences in interactions that are otherwise attributed to various factors towards technology, such as experience with robotics and demographics.
3) Attempts to scaffold trust through enacting social/social-like behaviours in the robot may even have a counter-intuitive negative impact on perceived trustworthiness in scenarios where users' trust towards the deployer is low.
4) Inclusion of the user in HRI scenario development promotes trust in the deployer and, indirectly, the robot.

We hope that future work exploring methods and metrics to measure trust in HRI takes this model into consideration.

ACKNOWLEDGMENT

REFERENCES

[1] R. E. Yagoda and D. J. Gillan, "You want me to trust a robot? the development of a human–robot interaction trust scale," *International Journal of Social Robotics*, vol. 4, no. 3, pp. 235–248, 2012.

[2] P. Hancock, D. Billings, K. Schaefer, J. Chen, E. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human factors*, vol. 53, no. 5, pp. 517–527, 2011.

[3] K. E. Schaefer, J. Y. Chen, J. L. Szalma, and P. A. Hancock, "A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems," *Human factors*, vol. 58, no. 3, pp. 377–400, 2016.

[4] T. Law and M. Scheutz, "Trust: Recent concepts and evaluations in human-robot interaction," in *Trust in Human-Robot Interaction*. Elsevier, 2021, pp. 27–57.

[5] K. Vicente, *The human factor: revolutionizing the way we live with technology*. Vintage Canada, 2004.

[6] D. Cameron, S. de Saille, E. C. Collins, J. M. Aitken, H. Cheung, A. Chua, E. J. Loh, and J. Law, "The effect of social-cognitive recovery strategies on likability, capability and trust in social robots," *Computers in Human Behavior*, vol. 114, p. 106561, 2021.

[7] D. Y. Geiskkovitch and J. E. Young, "Children's overtrust: Intentional use of robot errors to decrease trust," in *29th International Conference on Robot-Human Interactive Communication (SCRITA Workshop)*, 2020.

[8] R. C. R. Mota, D. Rea, A. Le Tran, J. Young, E. Sharlin, and M. Sousa, "Playing the 'trust game' with robots: Social strategies and experiences," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2016, pp. 519–524.

[9] L. D. Riek, "Wizard of oz studies in hri: a systematic review and new reporting guidelines," *Journal of Human-Robot Interaction*, vol. 1, no. 1, pp. 119–136, 2012.

[10] J. Trafton, L. Hiatt, A. Harrison, F. Tamborello, S. Khemlani, and A. Schultz, "Act-r/e: An embodied cognitive architecture for human-robot interaction," *Journal of Human-Robot Interaction*, vol. 2, no. 1, pp. 30–55, 2013.

[11] D. Cameron, J. Aitken, E. Collins, L. Boorman, A. Chua, S. Fernando, O. McAree, U. Martinez Hernandez, and J. Law, "Framing factors: The importance of context and the individual in understanding trust in human-robot interaction," in *IROS 2015: Workshop on designing and evaluating social robots for public settings*, 2015.