Framing Factors: The Importance of Context and the Individual in Understanding Trust in Human-Robot Interaction

David Cameron, Jonathan M. Aitken, Emily C. Collins, Luke Boorman, Adriel Chua, Samuel Fernando, Owen McAree, Uriel Martinez-Hernandez, James Law[1]

Abstract— **In this paper we explore the factors and methodologies from a range of disciplines used to investigate trust in human-robot interaction (HRI). Our investigation highlights a growing field, which recognises the importance of understanding the deployment of robots in real-world settings, but where a lack of common definitions and experimental clarity impedes the development of a comprehensive framework for investigation. As a result, we propose a bottom-up approach that emphasises context and user perspective as the foundation for future investigations into trust in HRI.**

I. Introduction

Robotics has been identified as an area with the potential to overcome challenges ranging from caring for society's ageing demographic [1], to enabling the next generation of advanced manufacturing [2]. As robots become more prevalent, particularly in these domains but also in public settings, there will be an increasing requirement for more natural human-robot interaction (HRI). An important aspect, particularly in light of common societal concerns about the safety of "intelligent" robots, is that such systems must be designed to engender trust in their users from the outset; both to encourage interaction, and reduce these fears.

Trust has been identified as foundational for successful interpersonal cooperation [3], and as a more general construct underwriting social order [4]. To enable us to create robots that can build trust in their use, we require a set of methodologies that enable us to measure and evaluate user trust in our creations. However, despite the ongoing effort to define 'trust' it remains a vague concept because the meaning of 'trust' is dependent first and foremost on the context in which it is being discussed as has been highlighted by, for example, Bauer [5] who frames a definition of trust around the expectation of one agent for another agent's behaviour to be particular within a certain situation. The issue within the context of HRI comes primarily from the application of methodologies around which trust can be explored and measured as a variable. Different agents in different contexts will necessarily have differing starting points for the level of trust they hold for a robot in an given scenario.

This work was supported by EPSRC Grant No EP/J011843/1 (Reconfigurable Autonomy), EU grant no. 612139 WYSIWYD "What You Say Is What You Did", and the European Union Seventh Framework Programme (FP7-ICT-2013-10) under grant agreement no. 611971.

[1]Sheffield Robotics, University of Sheffield, Pam Liversidge Building, Sir Frederick Mappin Building, Mappin Street, Sheffield, S1 3JD, United Kingdom d.s.cameron, jonathan.aitken, e.c.collins, l.boorman, dxachua1, s.fernando, o.mcaree, uriel.martinez, j.law@sheffield.ac.uk

It is, therefore, particularly important to consider methodologies and measures in terms of the context in which HRI occurs. For example, methodologies to explore measures of trust in the context of user-safety in robotic-assisted manufacturing [6], may differ substantially from those available to explore trust in a robot-led way-finding HRI scenario [7]. As a field at the intersection of many disciplines (e.g., engineering, computer science, psychology, sociology, etc.), there are many methodologies to be considered.

Hancock et al. [8] provide a solid starting point for research in this field by identifying a wide range of factors, spanning multiple disciplines, that influence trust in HRI. By examining existing studies into these factors we aim to identify a range of approaches, from a variety of fields, that when combined may provide an effective template for the ideal human-centred study into trust in HRI. However, as we begin to explore these ideas we discover that there is still much ambiguity in how these factors are defined, and that experiments often fail to isolate the effects of individual influencing factors. Our investigations highlight the importance of careful experimental design, and within that the importance of context when exploring factors impacting on trust.

This paper builds on the work by Hancock et al., by examining the variety of methodologies used to investigate these factors as reported in some of the key papers in these areas. Importantly, we find that the complexity of the subject prohibits the formulation of a comprehensive approach to studying trust in HRI, and highlights the need to maintain focus in investigating a single aspect of trust specific to the given context. We will demonstrate, with examples, the difficulty in using the term trust within HRI in a generic fashion by giving examples of experiments in which trust has been explored, but which cannot be clearly compared due to divergent contexts. This is not a paper about how to adequately define trust, but rather one which aspires to highlight the difficulties in exploring a conceptual variable (trust) in a way that allows for the results of different studies to ultimately be compared.

In the remainder of this paper we will review a selection of the most prominent examples of research methodologies applied to investigating trust in HRI (following Hancock et al.'s factors [8]), highlight the difficulties with categorising and isolating factors for investigation, and propose a framework for further research in this area that emphasises the subject and scenario as being fundamental to the experimental process.

II. Methodologies for studying trust

Hancock et al. [8] identify 33 factors influencing trust in HRI, grouped within 3 categories and 6 sub-categories. The main categories (and sub-categories) are: Human-related (ability-based, characteristics); Robot-related (performance-based, attribute-based); and Environmental (team collaboration, tasking). In Tables I-III we highlight a selection of these factors, and describe some of the most prominent works to have investigated them. In doing so we identify some common issues that make defining a comprehensive framework for investigating trust in HRI problematic.

Although these factors identify many different approaches to understanding antecedents of trust in HRI, study procedures tend to rely on the introduction of a fault or uncertainty in automated behaviour, and explore user response through either monitoring use of automation or surveying participants' trust in HRI. Multiple methods or use of converging measures are rarely used in these studies. This has implications for the confidence that the reader can ultimately have in the studies, because it is uncertain that the study manipulations chosen are solely influencing the intended studied factors (for further discussion, see Section IV). Moreover the studies reported by Hancock et al. all take place in lab-like conditions that, while likely well representing HRI in tightly controlled environments (e.g., manufacturing), offer little in the way of ecological validity. In short, the user is considered as another variable rather than the centre of a human-focussed HRI design.

In a truly human-focussed design, HRI studies should not just import constructs from social or cognitive psychology but also seek to best understand *how to explore them within the context of HRI*. Thus, HRI studies can be designed with the user as a central focus to accommodate user interactions in both naturalistic and theoretically meaningful situations. We propose that multiple methods or converging measures could be effectively used to approach this. For example, user-robot interpersonal distance, user self-reports of liking, and coding of user facial expressions in a single field study all converge to indicate individual differences in the impact of a humanoid robot's simulated facial expressions on users' liking of the interaction [29]. A solid evidence base and careful consideration of social psychological literature enables a human-focussed theoretical account of such findings to be developed and explored.

III. Reflection on the meta-analysis

Hancock et al. approach understanding engendering user trust during HRI by identifying critical related factors [8]. They broadly categorise factors for trust as being Human, Robot, or Environmental in their origin, and identify individual factors (many itemised above) that have demonstrable or potential influence on HRI. As the authors acknowledge, many of these factors remain to be formally explored in HRI scenarios. However, this meta-analysis [8] indicates an important issue beyond that of the dearth of empirical studies of trust during HRI. Their approach of understanding trust in HRI by itemising relative factors draws attention to the ambiguity involved in both defining and empirically exploring them.

Factors identified by Hancock et al. include those that are broad enough to encompass many others, which are listed alongside and identified as separate. For example, a robot's 'behaviour' is identified as a factor [p.523] but the robot's 'dependability', 'failure rates', and 'false alarms' are extracted as being factors independent from robot behaviour. This raises the questions of what constitutes a robot's behaviour, if these do not? And how, or at least in what effect direction, does robot 'behaviour' impact on trust?

Ambiguity within the factoral model is also found in the Human and Environmental themes. Within Human factors, items such as 'self-confidence' and 'demographics' are listed but without consideration for their exact impact on trust itself. Furthermore, factor ambiguity remarkably exists *across* themes: The Human factor 'operator workload', shares a substantial overlap with Environmental factors 'task complexity' and 'multi-tasking requirement', as both are used in psychological experiments to induce workload. A hypothetical future review based on this model is then left with the challenge of identifying which factor(s) experiments actually target.

In developing a series of isolated factors, the meta-analysis presents trust as a one-way relationship of users trusting a robot's behaviour [8]. However, psychological studies of human-human interactions indicate that trust is a dynamic and evolving process, rather than a fixed one [30]. Moreover, an individual's trust in a partner can develop through the process of being trusted by that partner, a phenomena termed reciprocal trust [31]. Given the broad and informative application of social psychological principles to HRI the study of reciprocal trust in HRI could offer substantial progress in understanding and fostering user trust. Exploring reciprocal trust requires robots to identify (or have pre-programmed recognition of) limits to their capacity to meet their goals. For example, mobile robots may encounter obstacles or barriers, requiring human intervention to allow the robot to progress [7], [32], [33].

In sum, Hancock et al.'s model demonstrates that exploring trust in HRI is a difficult endeavour. Factors can be hard to precisely define and, critically, isolate from others in an investigation. However, this work, in the growing field of HRI, argues well that there is no single factor that reliably impacts on trust in HRI. They further identify substantial gaps in the literature, and it still remains to precisely conduct experiments targeting many of these. They do identify areas which have reliable impacts on user trust but these are constructs built around the ambiguously identified factors and subsequently may best be viewed with caution. We address one example of the difficulty in isolating trust factors for experimentation, below.

IV. Case-study: The Importance of Context

We propose that the context of the use of a robot is important in defining the trust that is placed in it. This

TABLE I

HUMAN TRUST FACTORS

Factor (Subcategory)	Key Papers (Citations)	Methods for Investigating Factor
Attentional capacity/ engagement (Ability based)	[9] (140)	Parasuraman and Manzey [9] argue that less attentional resources directed to automation increases reliance on robots (trust). Their review paper collects simulated scenarios of user adoption of automation e.g., space station life support. Trust is measured in terms of uncritical acceptance of information coming from automated systems. They argue that less attentional resources can lead to complacency and reduction in detecting errors from automation.
Operator workload (Ability based)	[10] (20)	Desai et al. [10] report that lower cognitive load is associated with greater trust in robots. Their scenario required users to remote-control a semi-autonomous rescue robot around a maze; the robots reliability was manipulated across conditions as a means of nudging users to take control. Users could alter the degree of control they held over the robot and user trust was measured through established self-report surveys.
Prior experiences (Ability based)	[11] (704) [12] (520)	Lee and Moray [11], [12] discuss the level of experience a user has and how this impacts on their use of autonomy. Volunteers were tasked with operating a complex dynamic system - a simulated orange juice pasteurisation plant. Faults were injected into the system, which volunteers were required to recover from using manual, automated, or combined forms of control whilst simultaneously being tasked to log data. Volunteers were surveyed after their operational sessions. Results show that once the user became experienced with the system they could quickly adapt to the injection of faults, but they rapidly lost trust in the automated systems as their performance fell. This trust then required rebuilding as the users frame of experience with the autonomy was readjusted.
	[13] (258) [14] (1524)	Dzindolet et al. [13] investigate the performance of software to detect camouflaged soldiers within a photograph. Participants were rewarded for accurate detection and allowed to use either the autonomous detector (accuracy experimentally adjusted) or their own judgement. Users reported in surveys that they typically trusted their own experience over the detector, particularly if the aid malfunctions in ways that the operator cannot explain [13], although training on why errors arise can mitigate this effect.
Self-confidence (Characteristics)	[12] (520)	Lee and Moray's study of simulated operation of an orange juice pasteurisation plant [12] found that as user self-confidence in operation increased (measured through a self-report questionnaire), propensity to trust in automation decreased (measured through questionnaire and participant use of available automation). As described above (for the prior experiences factor) participants were assigned to varying conditions of plant reliability, introducing faults in either manual or automated control at pre-determined times.

TABLE II

ENVIRONMENTAL TRUST FACTORS

Factor (Subcategory)	Key Papers (Citations)	Methods for Investigating Factor
Communication (Team collaboration)	[15] (46)	It was found in [15] that when a robot's dialogue was adapted for expert knowledge (names of tools rather than explanations), expert participants found the robot to be more effective, more authoritative, and less patronising. This work suggests adaptation in human-robot interaction has consequences for both task performance and social cohesion. It also suggests that people may be more sensitive to social relations with robots when under task or time pressure.
	[16] (195)	Severinson-Eklundh et al. [16] reached the conclusion that for a service robot, addressing only the primary user in service robotics is unsatisfactory, and that the focus should be on the setting, activities, and social interactions of the group of people where the robot is to be used.

section examines previous investigations into trust factors, and highlights potential pitfalls in taking them out of context.

Lee and Moray [11], [12] and Gao and Lee [34] describe a valuable series of experiments conducted using a model of an orange juice pasteurisation plant. They use a software model through which the users can select an autonomous controller to operate the plant or undertake the procedures manually. These experiments identify and investigate a collection of factors that play important roles in the user trusting autonomous control systems. These include:

- Influence of load on the users - they are tasked to control the plant whilst being required to keep an accurate log.
- Influence of failures on the users - failures of plant equipment are randomly inserted, and the user can choose to use either autonomous or manual control to correct the problem. Additionally each control technique could fail to act as intended.
- The influence of feedback in the system - the operator is exposed to varying noise levels as a part of their decision making process. The operator then projects this onto the perceived reliability of the control systems.

This collection of experiments covers a wide range of factors identified as important in building trust in a system. However, each experiment contains a range of variables on the user that are not readily linked. For example, when the users are tested for the influence of workload, this factor is not isolated from the user and their personality; this is then immediately coupled with a failure which will silently increase user workload and their cognitive process. Varying levels of personal experience with automation will complicate the user's response, but are not adequately mitigated.

This highlights the difficulty in experimental design, where a series of factors are designed to be investigated within an experiment, but where a number of extra, silent, factors become included by accident. This modifies the response as each factor would influence choice, without being recorded. Figure 1 shows graphically how two distinct factors overlap and introduce ambiguity as experimental context weakens.

TABLE III
ROBOT TRUST FACTORS

Factor (Subcategory)	Key Papers (Citations)	Methods for Investigating Factor
Level of Automation (Performance-based)	[17] (1607)	Parasuraman et al. [17] define 10 levels of automation from low autonomy (level 1), where the human takes full control, to high autonomy (level 10), where the computer decides everything, ignoring anything a human does and providing no information. Selection of appropriate autonomy is important in engendering trust that a system, such as a robot, can complete a task with an appropriate level of human intervention, particularly if the nature of the autonomy will change how the operator will behave.
	[18] (194)	Goodrich et al. [18] experiment with how well these levels of autonomy relate to trust in a potential application. They designed a single human-robot system which had adjustable levels of autonomy which could be selected to complete a task. The human participant was given a secondary task, subtracting 7 from 3653, to complete which took their attention from controlling the robot. They found significant issues surrounding how tasks were terminated. They found that the level of autonomy could interfere with the decision making process of the operator confusing them when something happened they had not commanded. This caused more suspicion of the autonomous systems (especially under fault conditions). A similar effect has also been observed in an autopilot system by Parasuraman et al. [19].
	[19] (220)	
	[20] (132)	Ruff et al. [20] used a software simulation to provide participants in a study with control over a variable-sized fleet of Unmanned Aerial Vehicles (UAVs). These vehicles were programmed to complete various military missions autonomously, but could be interrupted by the supervisor. A decision-based tool could detect error states and notify the supervisor. It could take corrective action or allow the supervisor to correct the situation. Users abandoned the automation especially in complex situations where they were required to manage large numbers of UAVs, as they did not trust the autonomy to take corrective action.
Failure rates (Performance-based)	[21] (22)	Some studies have reported an increase in likeability and trust when errors were made [21] while others found that user-perceived errors had a negative impact on trust [22], [23]. In Salem et al. [22] volunteers were assessed on their compliance with a robot's unusual task requests as a behavioural measure of trust. Tasks included a breach of privacy (typing someone else's password into a laptop), odd reversible actions (throwing mail into the dustbin) or irreversible actions (pouring orange juice into a potted plant). Salem et al.[22] found that subjective measures of trust (self-reports of perceived trustworthiness) were independent their behavioural measure of trust (compliance).
	[22] (1)	
	[23] (14)	
Transparency (Performance-based)	[23] (14)	Desai et al. [23] investigated the effects of robot failures and transparency on trust. They examined trust in 'real-time', where participants were asked to rate their trust on the robot every 25 seconds while controlling the robot. They found that the post-run questionnaires were influenced by primacy-recency bias. Also, warning users of potential drops in reliability did not negatively influence trust during the interaction.
Proximity/co-location (Attribute-based)	[24] (78)	Bainbridge et al. [24] found that participants interacting with an embodied robot (co-located) had more trust in it than participants interacting with a video display (distant-located) of the same robot. They used 3 tasks as their measure of trust: a simple task (putting books on the bookshelf), an unusual task (placing books in the garbage) and a social task (amount of space given to the robot or screen when placing the books behind them). Participants were both more likely to comply with unusual tasks and walk up to the robot when the robot is physically present than in a video.
	[25] (3)	Haring et al. [25] further use user-robot proximity as a measure of trust: participants got closer to an android robot following repeated interactions with the robot, as they felt the robot was safer and less of a threat.
Robot Personality (Attribute-based)	[8] (88)	Robotic personality refers to the ability of a robot or personal computer (PC) to interact with people emotionally as well as on a logical level. It was noted in [8] that robot attributes such as personality had less impact than performance-based factors such as reliability. However [26] found that a serious, caring robot induced more compliance than a playful, enjoyable robot in a task where the robot assisted people with a series of breathing and stretching routines.
	[26] (89)	
Adaptability (Attribute-based)	[27] (40)	This has connections to the personality factor, above. It was found that when a robot adapts its personality to that of the human participant it is perceived to be more effective in assisting stroke patients with tasks [27]. The pilot experimental results provided evidence for the effectiveness of robot behaviour adaptation to user personality and performance: users (who were not stroke patients) both tended to prefer personality matched robot therapists, and performed more or longer trials under the personality matched and therapy style matched conditions.
Robot type (Attribute-based)	[28] (8)	Robot type was investigated by asking participants to rate the level of trust they felt when watching a video of someone passing different types of robots – the human controlled wheel chair was most trusted, while a large autonomous robot was least trusted [28]. This study demonstrates the multiple factors within "robot type" which moderate trust, e.g., both the size and method of control of the robot had a large impact on the level of trust reported.
	[25] (3)	Questionnaires and a trust game are used to assess the level of trust between humans and an android robot. The study was recognised by the authors as being limited in not having comparisons with other non-humanoid robots or humans.

Depending on the experiments chosen, their design and specification, the experimental cutting plane will move up or down. In the region of interference the influence of Factor A or Factor B can not be directly separated, producing confusion which actively influences the experiment.

The experimental cutting plane is strongly influenced by the selection, design, and setup of the experiment which is in turn defined by the context of the investigation. With a strongly specified system, the context of use can be isolated; each factor separately designed for and their interaction strongly controlled.

Studies such as [11], [12], [34], [13], [14], [18], [19], [20] rely on the operator taking decisions into whether to use an autonomous, or robot system, based on their trust in that system. The process for this decision making is underpinned by the user trust in the system, yet making that decision adds a cognitive load [35], an example of a silent factor, to the operation. This is especially true when failures are added into the mix: not only is the participant deciding whether to use automation or not, they will also be making decisions about the impact of failure mode on performance, changing the level of cognitive load during the experiments and exacerbating the silent factor.

To overcome this, the decision making process should

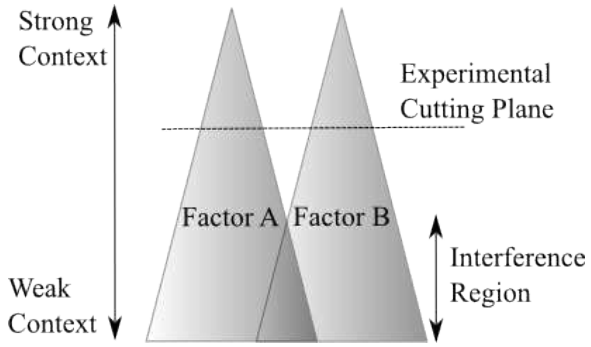

Fig. 1. Influence Cones with Context

be adequately explored to ensure that the experiment is targeted at factors without having extra silent augmentation. Techniques such as statecharts [36] play an important role in understanding the complex interplay of factors by providing a visual representation of the systems involved, allowing proper exploration and definition of the process. This allows the context to be strongly defined, and the factors understood and recorded correctly. This is especially important in any autonomous device where high levels of technology are linked to non-prescribed, complex human interaction (potentially with many different devices) [37].

This also stands when investigating trust for different systems. Here the context provides a vital role in separating the individual factors. It is important then that context is held as foremost in experimental design, as it is this that defines the optimum investigation. As typically seen in optimisation problems, there is "No-Free Lunch" [38], that is to say one technique will not solve all problems, but careful investigation of the problem, and the context, will reveal an appropriate technique.

V. Discussion and ways forward

An effective human-robot team requires some level of trust to be held between each of the agents involved: both the humans and robots. Acknowledging both the existence of these multiple perspectives, and that engendered trust will differ depending on which perspective is being taken, and crucially, on the context in which it takes place, is essential to the production of clear methodologies to explore trust in HRI. In a previous attempt to explore trust in HRI, Hancock et al. [8] employed a factors approach. This approach is systematic with its handling of the concept of trust, but in attempting to factorise this issue the model serves best to highlight the complexity of the problem of empirically tackling trust in HRI. We propose, alternatively, that empirical HRI studies desiring to focus on trust should primarily consider agent perspective and context of the human-robot interaction in order to begin to design an experiment that tackles trust.

Factors do exist which influence trust: the behaviour of those agents involved; where that behaviour stems from (culture, issues of communication, etc.); the type of task; and the physical environment in which the task takes place, to name just a few. However, these factors do not exist in isolation. They co-exist, and represent two – or often more – sides of the same issue (how does one truly separate culture and prior experience, or operator work-load and type of task?). Due to these factor's natures as co-existing concepts that influence trust, it is better to couch the study of trust in something more singular: the individual, and then subsequently the situation in which that individual is using a robotic tool. In sum, what we propose to solve this issue of the complexity of exploring trust within HRI is that researchers begin from the user – the hypothetical individual involved in the team – and work backwards from there to produce the most effective experiment in which to test the precise variable of interest by framing the whole endeavour around context from start to finish.

One way forward then, when thinking about designing the best methodology to test trust in HRI, is to acknowledge context from the outset. But what does this mean in practical terms? A robot is a tool to do a job, it does that job in a particular context, and the human or humans involved are the tool users. Thus at the beginning of any trust-focussed HRI experiment one of the first question asked should be, 'What kind of person is involved?'. Acknowledging the tool user will focus the experiment. The context comes from secondarily acknowledging what all the agents involved in the use of that tool - including the tool itself - are doing. Knowing this it is possible to then decide which agent's perspective is to be the focus of the experimental question. The interpretation of a factor involved in trust will differ depending upon the perspective of the agent involved. Clearly defining the context of the tool use and the area of particular interest for any one experiment will provide the necessary context from which to build a valid experiment. Once this is known, methods that converge on a factor can then be selected by how they would be predicted to affect an individual in a certain context, allowing for more precise analysis of any one given trust-based HRI scenario. The goal of researching trust within HRI should therefore be to disseminate this contextual way of thinking to the wider HRI community to increase progress in the field by accepting context driven experimentation, in place of seeking a single definition of trust to use in all cases.

VI. Concluding remarks

We have explored the factors and methodologies affecting trust in human-robot interaction spanning a range of disciplines. Our investigation highlights a lack of common definitions and experimental clarity, which prohibits the development of a comprehensive framework for investigation. As a result, we propose a bottom-up approach that emphasises context and user perspective as the foundation for future investigations into trust in HRI.

We appreciate that our approach describes an analysis of trust in HRI as best undertaken in small and precise steps. It is not an encompassing model that linearly lays out the construct. Rather, our approach reflects what needs to be a community endeavour; one that will grow in precision as members of the field begin to build concise, singular

experiments that can be brought together to give a clearer picture of the concept of trust in HRI as the field advances.

REFERENCES

[1] T. Mukai, S. Hirano, H. Nakashima, Y. Kato, Y. Sakaida, S. Guo, and S. Hosoe, "Development of a nursing-care assistant robot riba that can lift a human in its arms," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 5996–6001.

[2] European Commission, *FACTORIES OF THE FUTURE Multiannual roadmap for the contractual PPP under Horizon 2020*. Luxembourg: Publications Office of the European Union, 2013.

[3] D. J. McAllister, "Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations," *Academy of management journal*, vol. 38, no. 1, pp. 24–59, 1995.

[4] J. D. Lewis and A. Weigert, "Trust as a social reality," *Social Forces*, vol. 63, no. 4, pp. 967–985, 1985.

[5] P. C. Bauer, "Conceptualizing and measuring trust and trustworthiness," *Political Concepts-Committee on Concepts and Methods Working Paper Series*, no. 61, 2014.

[6] K. Eder, C. Harper, and U. Leonards, "Towards the safety of human-in-the-loop robotics: Challenges and opportunities for safety assurance of robotic co-workers'," in *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*. IEEE, 2014, pp. 660–665.

[7] D. Cameron, E. C. Collins, A. Chua, S. Fernando, O. McAree, U. Martinez-Hernandez, J. M. Aitken, L. Boorman, and J. Law, "Help! i cant reach the buttons: Facilitating helping behaviors towards robots," in *Biomimetic and Biohybrid Systems, Living Machines 2015*, ser. Lecture Notes in Computer Science, 2015, vol. 9222, pp. 354–358.

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

[9] R. Parasuraman and D. H. Manzey, "Complacency and bias in human use of automation: An attentional integration," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 52, no. 3, pp. 381–410, 2010.

[10] M. Desai, M. Medvedev, M. Vázquez, S. McSheehy, S. Gadea-Omelchenko, C. Bruggeman, A. Steinfeld, and H. Yanco, "Effects of changing reliability on trust of robot systems," in *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*. IEEE, 2012, pp. 73–80.

[11] J. D. Lee and N. Moray, "Trust, control strategies and allocation of function in human-machine systems," *Ergonomics*, vol. 35, no. 10, pp. 1243–1270, 1992.

[12] ——, "Trust, self-confidence, and operators' adaptation to automation," *International journal of human-computer studies*, vol. 40, no. 1, pp. 153–184, 1994.

[13] M. T. Dzindolet, S. A. Peterson, R. A. Pomranky, L. G. Pierce, and H. P. Beck, "The role of trust in automation reliance," *International Journal of Human-Computer Studies*, vol. 58, no. 6, pp. 697–718, 2003.

[14] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 39, no. 2, pp. 230–253, 1997.

[15] C. Torrey, A. Powers, M. Marge, S. R. Fussell, and S. Kiesler, "Effects of adaptive robot dialogue on information exchange and social relations," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 2006, pp. 126–133.

[16] K. Severinson-Eklundh, A. Green, and H. Hüttenrauch, "Social and collaborative aspects of interaction with a service robot," *Robotics and Autonomous systems*, vol. 42, no. 3, pp. 223–234, 2003.

[17] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 30, no. 3, pp. 286–297, 2000.

[18] M. A. Goodrich, J. W. Crandall, and T. J. Palmer, "Experiments in adjustable autonomy," in *Proceedings of IJCAI Workshop on Autonomy, Delegation and Control: Interacting with Intelligent Agents*, 2001.

[19] R. Parasuraman, M. Mouloua, and R. Molloy, "Effects of adaptive task allocation on monitoring of automated systems," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 38, no. 4, pp. 665–679, 1996.

[20] H. A. Ruff, S. Narayanan, and M. H. Draper, "Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles," *Presence: Teleoperators and virtual environments*, vol. 11, no. 4, pp. 335–351, 2002.

[21] M. Salem, F. Eyssel, K. Rohlfing, S. Kopp, and F. Joublin, "To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability," *International Journal of Social Robotics*, vol. 5, no. 3, pp. 313–323, 2013.

[22] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, "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*. ACM, 2015, pp. 141–148.

[23] M. Desai, P. Kaniarasu, M. Medvedev, A. Steinfeld, and H. Yanco, "Impact of robot failures and feedback on real-time trust," in *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction*, 2013, pp. 251–258.

[24] W. Bainbridge, J. Hart, E. S. Kim, B. Scassellati *et al.*, "The effect of presence on human-robot interaction," in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, 2008, pp. 701–706.

[25] K. S. Haring, Y. Matsumoto, and K. Watanabe, "How do people perceive and trust a lifelike robot," in *Proceedings of the World Congress on Engineering and Computer Science*, vol. 1, 2013.

[26] J. Goetz and S. Kiesler, "Cooperation with a robotic assistant," in *CHI'02 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2002, pp. 578–579.

[27] A. Tapus and M. J. Mataric, "Socially assistive robots: The link between personality, empathy, physiological signals, and task performance." in *AAAI Spring Symposium: Emotion, Personality, and Social Behavior*, 2008, pp. 133–140.

[28] K. M. Tsui, M. Desai, and H. A. Yanco, "Considering the bystander's perspective for indirect human-robot interaction," in *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, 2010, pp. 129–130.

[29] D. Cameron, S. Fernando, E. Collins, A. Millings, R. Moore, A. Sharkey, V. Evers, and T. Prescott, "Presence of life-like robot expressions influences childrens enjoyment of human-robot interactions in the field," in *Fourth International Symposium on New Frontiers in Human-Robot Interaction*.

[30] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "An integrative model of organizational trust," *Academy of management review*, vol. 20, no. 3, pp. 709–734, 1995.

[31] M. A. Serva, M. A. Fuller, and R. C. Mayer, "The reciprocal nature of trust: A longitudinal study of interacting teams," *Journal of organizational behavior*, vol. 26, no. 6, pp. 625–648, 2005.

[32] J. Law, J. M. Aitken, L. Boorman, D. Cameron, A. Chua, E. C. Collins, S. Fernando, U. Martinez-Hernandez, and O. McAree, "Robo-guide: Towards safe, reliable, trustworthy, and natural behaviours in robotic assistants," in *Towards Autonomous Robotic Systems (TAROS) 2015*, ser. Lecture Notes in Computer Science, 2015, vol. 9287, pp. 149–154.

[33] O. McAree, J. M. Aitken, L. Boorman, D. Cameron, A. Chua, E. C. Collins, S. Fernando, J. Law, and U. Martinez-Hernandez, "Floor determination in the operation of a lift by a mobile guide robot," in *Proceedings of the European Conference on Mobile Robotics (In press)*, 2015.

[34] J. Gao and J. D. Lee, "Extending the decision field theory to model operators' reliance on automation in supervisory control situations," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 36, no. 5, pp. 943–959, 2006.

[35] L. J. Skitka, K. Mosier, and M. D. Burdick, "Accountability and automation bias," *International Journal of Human-Computer Studies*, vol. 52, no. 4, pp. 701–717, 2000.

[36] D. Harel, "Statecharts: A visual formalism for complex systems," *Science of Computer Programming*, vol. 8, no. 3, pp. 231–274, 1987.

[37] G. H. Walker, N. A. Stanton, D. P. Jenkins, and P. M. Salmon, "From telephones to iphones: Applying systems thinking to networked, interoperable products," *Applied Ergonomics*, vol. 40, no. 2, pp. 206–215, 2009.

[38] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.