Escape from the factory of the robot monsters: agents of change

Richards, D

Original citation & hyperlink:
Richards, D 2017, 'Escape from the factory of the robot monsters: agents of change' *Team Performance Management: An International Journal*, vol 23, no. 1/2, pp. 96-108
https://dx.doi.org/10.1108/TPM-10-2015-0052

DOI 10.1108/TPM-10-2015-0052
ISSN 1352-7592

Publisher: Emerald

Team Performance Management: An International Journal

Escape from the factory of the robot monsters: agents of change

Dale Richards

Escape from the factory of the robot monsters – agents of change

Introduction

Increasingly media stories warn of the impending prospect that advanced technology is not just slowly entering our everyday working lives, but it may also be viewed as the harbinger of a process that finds humans being slowly replaced in some jobs (e.g. BBC - 'Will a robot take your job?'[1]). In many instances we see the technology advance into modern working environments, but it is generally adopted more within industries that already incorporate machines and advanced automation. The integration of such technology is normally associated with cost savings, faster production rates, and increased safety. Each of these key factors would normally fall on the human worker to ensure goals were met in relation to the objectives set for that particular job.

The introduction of automation is seen as a benefit in terms of allowing the human worker to be withdrawn from tasks that may be viewed as laborious, dangerous, difficult or dull. However, rather than completely replacing the human element within the work environment a disruptive technology may be introduced that enhances not only the workforce's capability, but augments the physical composition of the human team. This may be introduced in several ways ranging from advanced automation of existing processes, the insertion of autonomous software agents to assist with decision support, or even a combination of the two whereby a physical robot can either operate automatically or autonomously (Richards, Stedmon, Shaikh & Davies, 2014). Indeed, we are seeing an increase in advanced robotics whereby the robot is composed of an agent-based model (ABM) that allows the individual robot to be connected not only to other robots, but to a much wider network. Thus allowing the collective team to make decisions based on a far richer understanding of

[1]http://www.bbc.co.uk/news/technology-34066941

pertinent information that is available. Both automatic and autonomous solutions to completing tasks raises important questions associated with the dynamics of the team.

The growing trend of introducing service robotics into the workplace is clearly evident. In 2014, robot sales across the world increased by 29% to 229,261 units in comparison to the previous year[2]. Although the majority of these figures are simply indicative of the increasing use of service robotics in major industries such as automotive and electrical, there is still an evident need as more companies turn to robot solutions in order to maximise production and promote business growth. The modern factory is quickly aligning itself to the cultural change forced upon it by the advances in technology that lends itself to the nature of tasks traditionally carried out by human workers.

This cultural phenomenon, where we do not raise so much as an eyebrow when we encounter solutions that involve some form of advanced robotics or agent-based system is becoming more pervasive in its nature. It is not just the realms of Science Fiction that displays the use of robotic systems assisting humans, but we see the use of human-robot teams in advanced space systems (Singer & Akin, 2010). Even day to day activities such as a simple trip to the museum that can result in encountering a robot tour guide (Shiomi, Kanda, Ishiguro & Hagita, 2006), or on arrival at a hospital can expose us to the use of robot helpers (Thiel, Habe & Block, 2009). We are even promised the use of robot helpers that can assist our everyday tasks (e.g. Kim, Cha, Park, Lee & You, 2011) or to support the elderly to continue remain living in their own home (Prakash, Beer, Deyle, Smarr, Chen, Mitzner, Kemp & Rogers, 2013). However, as the growing presence of robotic systems become more ubiquitous a number of interesting questions begin to surface in relation to how humans not only interact with robot systems, but the very nature of how teams can be composed of both humans and robots and the consequent dynamic this entails.

[2]International Federation of Robotics. http://www.ifr.org/industrial-robots/statistics/ Accessed on 20th October 2016.

This paper will initially explore the nature of team composition as defined by research focussed on teams composed of humans and the roles they adopt to accomplish goals. It will then examine this in the context of how humans interact with robots and indeed how human and robot members may interact within the same team. In order to understand the nature of such a complex dynamic, this paper will discuss different frameworks of control and delegation that will allow both human and robot members to achieve individual or shared goals.

Teams

There have been many instances where we have witnessed the importance of effective team performance in relation to events such as natural disasters, terrorist attacks and complex surgical procedures. In all of these examples the outcome is somewhat dependent on an effective team performing well as a collective whole. Early research in this area suggested that approaches to studying how humans behave and perform in teams was fragmented and received little attention in terms of theoretical understanding (Dyer, 1984). However, since that time it has gained more attention and study, resulting in several different theoretical explanations of the factors and dynamics involved in such a complex process (Salas, Stagl, Burke, & Goodwin, 2007). Salas, Cooke & Gorman (2010) state that examining the nature of team constructs requires a multi-disciplinary approach in order to gain a better understanding of collaborative and coordinated behaviours.

There is an intuitive feeling that given a number of team members it would be likely that they would achieve their goals more effectively than the same number of individuals who are not working together as a team. Gigone & Hastie (1997) suggest that this is indeed the case, but warn that the result (or output) may not always be superior to individual outputs. This is particularly relevant when comparing the sort of goal that has been set (i.e. a task that requires a quantitative output versus a creative one or one that involves complex decision-making). Of course, the context within which the task is realised is critical in determining whether a team approach or individual approach is bestsuited to the challenge. Not surprisingly, considering the interest in how military teams functioned, there were a number of extensive studies carried out since the 1940s (Benne & Sheats, 1948; Bales, 1950). It was Belbin's (1981) seminal work that described the benefits of good team composition and the importance on the team dynamic. Belbin defined a number of roles assigned to team members in order to empirically evaluate the properties of each team member, and he later developed a metric to assess individual team roles - the Self-Perception Inventory (SPI). Following several iterations, Belbin (1993) defined several key roles that are normally adopted within a team (see Table 1).

Role	Associated Behaviour
Implementer (Im)	Translates and applies the top-level concepts/plans
Co-ordinator (Co)	Organises, co-ordinates an controls the activities of the team
Shaper (Sh)	Challenges the norm, acts as devil's advocate, assists in motivation and winning
Plant (Pl)	Puts forward new ideas/strategies to achieve the goal
Resource Investigator (RI)	Focuses outside the immediate team for other ideas and resources that can assist in achieving the goal
Monitor Evaluator (ME)	Assists in analysing and evaluating the different ideas originating from within the team against achieving the goal
Team Worker (TW)	Ensures the team members stay together fostering team spirit and providing support
Completer/Finisher (CF)	Oversees the standards of tasks within the team and errors are kept to a minimum. Also provides focus for keeping to time in achieving the goal

Table 1 – Belbin's definition of team roles

Belbin's definition of team roles highlights the many different roles associated within a team that provide the key elements for ensuring an effective team, with dynamic role allocations being distributed amongst members that increase the likelihood of achieving the team goal. If we were to translate this approach in the context of a team that included robots (or software agents), where membership of the team is composed of both humans *and* software agents, then we can begin to focus on roles that are more likely to be associated with either human or agent team members. For example, the roles of Implementer and Co-ordinator can be defined as belonging to the leader of the team; as it is the Implementer that sets the goal to be achieved and the Co-ordinator who allocates the task(s) to each of the members within the team. However, elements of all the behaviours as defined by Belbin can be discussed in terms of different characteristics which may be distributed across a team, and in essence could be attributed and *shared* between both humans and robots.

Defining and allocating roles is not only a method used to facilitate effective team behaviour but is also indicative of the importance assigned to communication between different team members. Communication is a critical aspect to consider in terms of coordination and sharing of information within a team (Cooke, Kiekel, & Helm, 2001). If there is a failure or degradation of communication then it is likely to expect reduced team efficiency and a lower likelihood of that team in achieving their goal. By combining the hierarchical roles within the team and then examining the importance of establishing and maintaining effective communication, we can begin to view a functioning team as a holistic structure that utilises processes in order to achieve its goal. A team may also possess processes that drive the mechanics of the team that may be directly associated with individual team behaviours, or equally they may be more strategic in nature. These range from the nature of communication between team members, co-ordination, etc. Although other team factors may affect the nature of team processes (such as the size of the team, how new the team is, etc), such processes are critical to contributing to team success. These processes are guided by implicit social norms that all members are expected to align, and in essence act as regulators of expected behaviours within the team and assist with identity formation within the team (Feldman, 1984). The alignment of team members to an exemplar norm is important when considering human members within the team, and equally their perception of non-human team members.

Human-Robot Interaction

Previous research has examined the nature of human-robot interaction (HRI) between humans and individual robots, or indeed smaller teams of robots (Murphy & Rogers, 2001; Schultz & Parker, 2002). However, we know little of the integration of agent-based systems as members within a human-agent team. The Defence Advanced Research Project Agency (DARPA) programs referred to as MICA (Mixed-Initiative Control of Automa-teams) and SHARC (Stochastic, Hierarchical, Adaptive, Real-Time Control) both investigated human-agent interaction by developing architectures that allowed the supervision of multiple unmanned systems by a small number of human operators. Linegang et al. (2003) reported the importance of sharing the capabilities of the system (in terms of the "automation") with the human, and the ability for the operator to interact with the system developed under MICA and SHARC. The ability of the human to understand what the system is doing (and its intent) remains a key design requirement for human-robot teams, even when the human operator is untrained in the use of the system (Nevatia et al. 2008).

In order to understand what the non-human element of a team is doing it is crucial that the human team members are able to ascertain information about state and intent of the non-human agent. This is more critical when we consider teams of robots operating within a bounded ABM construct, allowing the non-human elements of a team to share information and decision-making as a collective. In human teams we tend to rely not only on verbal protocols to establish and perceive states, but non-verbal communication also. Previous studies have examined the sort of cues required by humans and found that humans perceive robots in terms of not only their physical appearance, but also their perceived abilities and performance (Adams et al, 2003; Goetz et al, 2003; Schermerhorn & Scheutz, 2011). In some instances it has been argued that the best way of achieving a higher level of interaction between a human and robot is for the robot to mimic (to some extent or another) the behaviours of their human counterparts. For example, Trafton et al (2013) proposed a cognitive architecture that would capture human interaction and then apply it to human-agent behaviours. This may also be applied to more complex cognitive behaviours that humans display and then translated into a cognitive architecture that takes human-robot interaction into account (Holland et al, 2013). Of course, the extent to which a robot system can utilise such a cognitive architecture is limited to the sensor fit associated with the individual robot. For example, if behaviour between human team members recognises a change in how information is exchanged between members based on physical proximity, then this behaviour could only be replicated by robots if they possess the ability to sense other team members within their vicinity.

While there are clear differences between human and robot members of a team, we need to understand the boundaries that exist between human-robot capability, beliefs, intent and control. It is only through this we can begin to design an effective human-robot team that can work effectively together. To a large extent we can view the use of non-human agents as extensions of ourselves, in that the human may delegate authority to an agent to perform a number of tasks that allow a goal to be achieved. However, this perception only works if the human can precisely define the task to be delegated and then ensure the task is tightly bound in relation to the potential outcomes. For example, if we chose for a robot to achieve Goal A, then they must be able to sequentially complete task A, B and C. However, the robot is bound to the law that they may only perform one task when the condition is met that the preceding task is completed. Again, this tightly coupled to team roles and associated processes.

Human-Robot Teams

There are many ways by which a human may interact with an agent-based system, and these vary dependent on the nature of the system being developed. Scholtz (2003) suggests that allocating roles is perhaps more critical in relation to the human-robot team, knowing whether individuals (human or agent) are assigned as (1) supervisor, (2) operator, (3) teammate, (4) mechanic/programmer, or (5) bystander.

Previous studies have stated that the key to an effective HRI is the increase in robot performance and the decrease in human input (e.g. Kaupp & Makarenko, 2008), therefore suggesting that the role of the human within human-robot teams can be granulated to the form of physical command inputs. While it is plausible to assume that this may be a measure of good interaction design (i.e. to the detailed layout of the human machine interface) it does not account for the information or shared situation awareness required to allow the human to comfortably monitor the non-human element within the team. This would suggest that the native cognitive functions need far more consideration in terms of how tasks are allocated/shared and ultimately the process by which decisions are made.

With the integration of human and agent team members it is critical to understand the nature of decision-making and how the boundaries of such actions are shared and/or delegated between the two different team members. In order to understand the team composition within which decisions are made we must appreciate that there are different types of human-robot team composition, some of which are outlined in Fig 1.

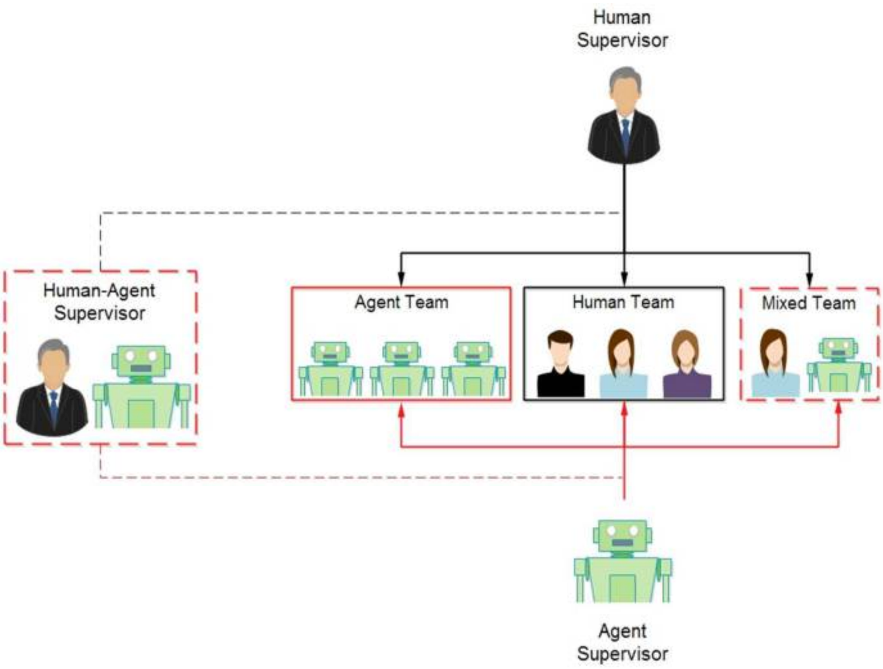

Figure 1 - Human-Agent teaming

This proposes several potential ways in which agents may be incorporated into a team. In the majority of instances we tend to think that a human supervisor would set a top level goal and then distribute that down to the team members. The supervisor would then monitor and interact with the team in order to achieve that goal. The alternate model would dictate an agent supervisor potentially generating and setting goals to the team. An agent supervisor of a human team however raises a number of issues in itself. The use of agents to assist, or partner, a human supervisor would perhaps be more acceptable as the interface between the supervisor and the team is human (regardless of how the decision or plan was calculated). However, if the agent becomes the supervisor the only members that would not question or interrogate tasks would be agent members of the team (unless they had been designed to do so). This brings us to a key element in any team interaction - trust. The key quality of trust relies solely on the nature of interaction between the agent system and the human (White & Richards, 2006) and if the agent performs the task as expected by the human (Hancock et al, 2011). Not surprisingly trust is a key human factor within human teams also (Dirks & Ferrin, 2001).

As we begin to consider the heterogeneous mix of human and agent team members we can begin to think of roles as being either fixed to an individual, shared, or dynamically changing over time. For example, in Figure 2, we can see how a goal may be set by the human supervisor and then passed onto the human-agent team. As this process develops within the team the role of supervisor shifts from the goal originator to the members of the team. This in turn allows the team to designate a sub-supervisor that will monitor the progress of the tasks, as the human and/or agent performs the behaviour associated with that particular task. This allows for the dynamic allocation of roles and related behaviours within the collaborative team. However, there would be a number of foreseeable instances that would create real or artificial boundaries that would preclude some team members from fulfilling a role (or task). For example, safety critical systems may employ agent behaviours to perform laborious monitoring of a system state, but would ensure that a human team member was responsible for overseeing that behaviour. Alternatively an agent may perform a series of complex behaviours, but stop short of a final action unless authorised by the human. However, in other instances, perhaps where safety was not a factor, then an agent could monitor and adjust behaviours to ensure an optimised effect.

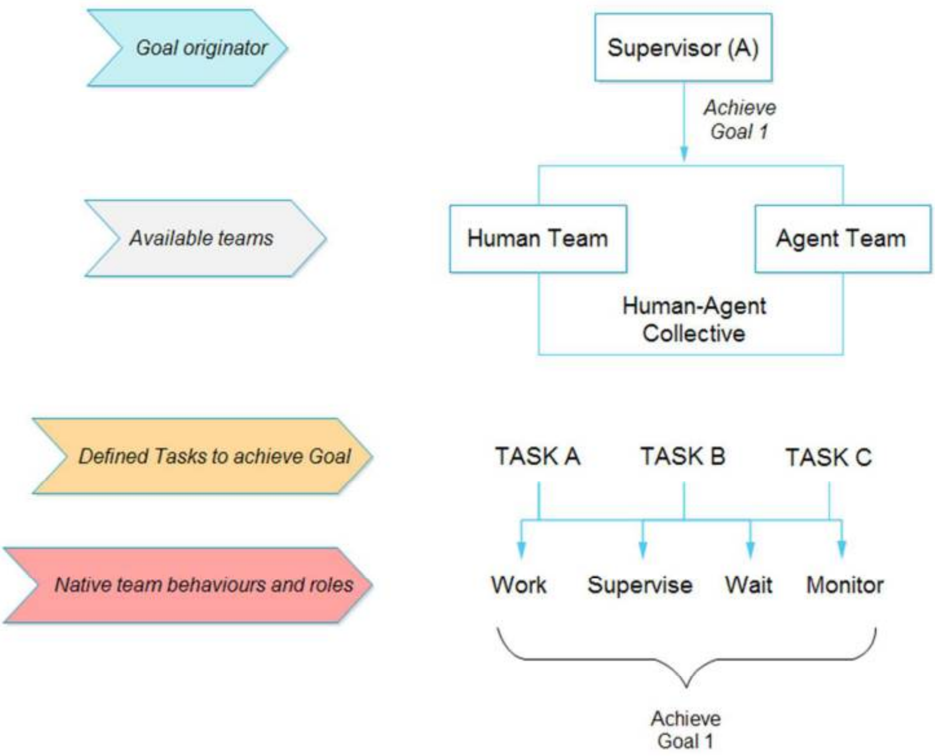

Figure 2 - Outline of human-agent team behaviour and roles

The cohesion between team members can determine the level of membership each individual has to the team, and ultimately the extent to this can determine whether the team achieves their goal (Mullen & Cooper, 1994). Goodman et al (1987) found that teams that were higher in cohesive behaviours possessed more power over their team members when compared to lower cohesion teams. However, conflicts are more likely to occur in highly cohesive teams that have tasks or goals that are at odds with one another (Thomas, 1992).

Goals

When we talk about teams it is important to discuss the rationale for why a team has been assembled in terms of its purpose and function. The composition and membership of that team serves to facilitate the achievement of the goal that is defined and set as a target that the team must achieve. There are several frameworks that exist within computer science that allow us to explore how complex agent-based systems can co-operate in order to achieve a goal. For example, the theories of Joint Intentions (Levesque, Cohen & Nunes, 1990), Shared Plans (Grosz & Sidner, 1990), Collective Intentions (Dunin-Keplicz & Verbrugge, 2002), Norm Internalisation (Andrighetto, Villatoro & Conte, 2010) and Collaborative Discourse (Rich, Sidner & Lesh, 2001).

Each of these approaches define ways in which behaviours are constrained and bounded, based on context and information available to the agent. However, with intelligent systems we can reasonably assume that the agent, after being given a goal, will use a degree of freedom in order to generate their own plans, actions, and beliefs as they strive to achieve their goal. It can be suggested that this is the same level of bounded rationality that exists within human teams. By its nature, the agent member belonging to an ABM tends to be defined by its reliance on the information that is either directly fed to it (by the human, other systems, other agents, or its own sensors). This is very much how we would also define a human team member, in that they operate solely upon the information with which they are provided.

When we think of how we define a goal then it is not only the context that that is important, but the nature of how such knowledge and decision-making is represented between the human and the agent. The goal must be defined in terms of the agent representation of the world in order for it to be processed appropriately, i.e. it is pointless telling an agent to locate a specific item if it does not have the appropriate means by which to search and identify the target. Similarly, we can look at human mental models and how such internal constructs about the world can assist the processing of information in order to achieve a goal (Johnson-Laird, 1983, 2006). The creation of any model, albeit human or agent, is determined by how information is processed and constructed in order to create structures of knowledge that represent the beliefs and states of an individual person or system. There is, however, still some discussion as to whether the nature of reasoning associated with such models is rule-based (O'Brien, 2009), determined by probability (Gigerenzer, Hoffrage & Kleinboltin, 1991), or based on inferences (Rips, 1994; Braine & O'Brien, 1998).

When we consider a mixed human-agent team we must consider the nature of how each of those team members internally represent knowledge and, more importantly, act on such information. If we assume that both human and agent members possess bounded rationality then the ability to share not only the goal, but the way they analyse (and act upon) information is a key component within the team process. Baxter & Richards (2010) discuss the importance of better understanding the nature of such model representation when sharing goals between both human and multiple agent systems. They further define the differences between different types of goals that may exist within a human-agent partnership, see Table 2.

Type of Goal	Description
User specified	This is a command sent by the human that reflects a top level goal. Several tasks may then be constructed in order to achieve this goal.
User delegated	This is a goal that has already been set by the human, but the agent system may attempt to achieve the goal, but will be required to have a level of human interaction to achieve the goal.
Internal system	These are purely goals generated by the agents in order to achieve the top level goal as set by the human. The agent system may therefore generate multiple local goals in order to complete tasks that will bring it closer to the user specified goal. The human may not be aware of the internal system goals generated by the agents.

Table 2 – Goal definitions and ownership

This allows us to define goals in relation to who sets them and the constraints associated with achieving that goal.

Frameworks of Human-Agent Autonomy

The way we interact with an ABM is very much determined by the framework of control that allows us to delegate the degrees of *authority* between human and machine. There are many ways in which the human can define this dialogue of tasking between human and agent. Both Sheridan & Verplank (1978) and Parasuraman et al (2000) have proposed a framework of authority by which a human-machine co-operative task may be shared or delegated. The nature of any framework of delegation of control will dictate the level of interaction between the human and the machine. In some instances it has been suggested that the control of tasks can at times be *traded* between both human and agent; defined as a level of adaptive control (Sheridan, 1992; Sarter & Woods, 1997). Sheridan (2011) discusses how tasks can be performed through control loops which may be considered as inner or outer control loops; whereby control is either through direct control (inner) or simply monitored by the controller (outer). The controller in this instance may be human or an autonomous agent acting on behalf of the human (and the associated constraints this brings with it). We must also view the state of being within the inner and outer two control loop as not being a binary condition; in that instances may very well dictate an individual (human or agent) to traverse between both inner and outer control states. This would imply a flexible framework of automation and/or autonomy that would allow variable levels of control to exist between both human and machine. In essence an adaptive (or variable) framework of control that requires a degree of delegation between both human and agent elements within any system. There are a number of different models that outline a number of frameworks of control whereby a human may delegate varying levels of authority with an autonomous system (Richards & Stedmon, 2015).

The use of an adaptive and intelligent agent team member may also be viewed as possessing the ability to traverse different functional areas within the organisation in order to obtain a specific goal. Webber (2002) referred to such dynamics as cross-functional teams. Daspit et al. (2013) suggest that shared leadership and cohesion are the most important aspects of teams when considering CFT. Members are more likely to participate in aspects of shared leadership when they perceive to have a shared purpose associated with a defined goal. Kirkman & Rosen (1999) highlight the importance of shared goals and a common purpose as being associated with positive influences within the team.

What we may take away from this is that any human-agent team will require a control framework upon which the nature of interaction is defined and set by firm rules. This formal interaction paradigm acts as a force in bounding rationality between the human and agent elements within a team.

Discussion

It is evident that advanced systems, whether they take the form of robotics or software agents, are becoming more common-place alongside existing human workers. In some instances the agent system is integrated into the team as a slaved system, providing automated outputs that are predictable and allow the user to be able to understand the system's intent. The introduction of agent-based autonomy within the team not only offers an increase in robot capability, but also an opportunity to integrate agent and humans within the same team. There are however different ways in which this team composition may be realised, with many roles within it needing to be fulfilled (Belbin, 1993). A human-agent collective will need to possess a degree of flexibility in order to not only share goals, but also allow for the dynamic delegation of authority between human and agent. In order for this to take place a formal framework of control would need to be in place; thus forming a structured control architecture that would allow interaction between both human and agent. And in some instances this could even be viewed as an attempt to achieve an alignment between the

ABM and human mental model in terms of sharing tasks/goals. In some instances the control of a given task may also *switch* between human and agent team members, thus making the system more flexible and responsive to changes within the context of the desired goal. To this end we may view the goal as being shared between both agent and human team members, with successful achievement of the goal acting as the catalyst.

In order for this team dynamics to take place, we must acknowledge that a number of pertinent human factors issues play an important role in whether the human-agent partnership is effective. Each individual team member will possess their own unique understanding as to their role and what they will need to do in order to achieve the goal allocated to them. This applies to both human and agent-based systems, although the human will perceive the non-human element within the team bottom-up (in that it is merely a simple machine that is slaved to the human), or alternatively adopt a top-down processing perspective (perceiving the agent as an equal member of the team). A top-down approach would allow the dynamic of the team to form along the same lines as traditional human teams, with defined roles, behaviours, communication, norms and processes. In any instance an effective team will need to achieve what Mathieu et al (2000) referred to as a 'convergence' of mental models between team members in order to achieve a highly effective state (and outcome).

It is important to stress that a multidisciplinary approach is required to examine team performance, due to the different factors that come into play when both humans and machines interact. A single metric (or protocol) has been put forward that can be used to measure the effectiveness of teams that can be validated. However, with the increase of robots being integrated into everyday procedures some methodologies have shown promising ways in which human-robot team members can be monitored and assessed. For example, Tiferes et al. (2016) examined the integration of surgical robots within surgery teams whilst using different methods of recording visual/audio data in order to assess team communication and interaction. This would highlight the need of obtaining such data in terms of considering whether teamwork in these particular instances is effective.

However, it is important to adopt subjective assessment techniques already being used within Social and Organisational Psychology to ascertain how human co-workers interact and perceive robotic team members. In order to assess this, a multidisciplinary approach is needed that will assist in our understanding of not only the quantitative value of integrating robot systems alongside human team members, but the effect this has on the existing human team in terms of the cognitive impacts (such as trust and attitude). By focussing solely on quantitative measurements it is possible to arrive at an incomplete answer. Yes, productivity may be seen to rise initially, but closer examination would be needed to identify the nature of human-robot team interaction, especially in terms of the potential for introducing an increased likelihood in error (e.g. due to a lack of robot intent being displayed).

The equality we view an ABM may also afford us to suggest that the agent may be regarded as a social agent in many regards in that it shares not only the goal that the team all strive towards, but also prepares the foundation for a human-agent collective whereby a single team is perceived that possess a *shared bounded rationality*.

References

Adams, B., Bruyn, L., Houde, S. & Angelopoulos, P. (2003) Trust in automated systems: literature review. Defence Research and Development Canada Toronto No. CR-2003-096, 2003.

Andrighetto, G., Villatoro, D., & Conte, R. (2010) Norm internalisation in artificial societies. AI Communications, 23(4), 325-339.

Bales, R. F. (1950). *Interaction process analysis*. Cambridge, MA: Addison-Wesley.

Baxter, J. & Richards, D. (2010) Whose goal is it anyway? User interaction in an autonomous system. Association for the Advancement of Artificial Intelligence Conference: Atlanta, USA.

Belbin, R. M. (1981). Management Teams: Why They Succeed or Fail. Oxford: Butterworth-Heinemann.

Belbin, R.M. (1993) Team Roles at Work. Butterworth-Heinemann, Oxford.

Benne, K. D., & Sheats, P. (1948) Functional roles of group members. Journal of Social Issues, 4, 41-49.

Braine, M.D.S., & O'Brien, D.P. (1998). Mental Logic. Mahwah, NJ: Lawrence Erlbaum Associates.

Cooke, N. J., Kiekel, P. A., & Helm, E. E. (2001). Comparing and validating measures of team knowledge. In *Proceedings of the Human Factors and Ergonomics Society,* 361-365.

Daspit, J., Tilman, C.J., Boyd, N. & Mckee, V. (2013) Cross-functional team effectiveness: An examination of internal team environment, shared leadership, and cohesion influences. Team Performance Management: An International Journal, 19(1), 34-56.

Dirks, K.T. & Ferrin,D.L. (2001) Trust in leadership: Meta-analytic findings and implications for research and practice. Journal of Applied Psychology, 87(4), 611-628.

Dunin-Keplicz, B. & Verbrugge, R. (2002) Collective intentions. Fundamenta Informaticae, 51(3), 271-295.

Dyer, L. (1984) Studying human resource strategy: An approach and an agenda. Industrial Relations: A Journal of Economy and Society, 23(2), 156-169.

Feldman, D.C. (1984) The development and enforcement of group norms. Academic Management Review, 1, 47-53.

Gigerenzer, G., Hoffrage, U. & Kleinbölting, H. (1991) Probabilistic mental models: A Brunswikian theory of confidence. Psychological Review, 98(4), 506-528.

Gigone, D. & Hastie, R. (1997) Proper Analysis of the Accuracy of Group Judgments. Psychological Bulletin. 121(1):149-167.

Goetz, J., Kiesler, s., & Powers, A. (2003) Matching robot appearance and behaviour to tasks to improve human-robot cooperation. In Proceedings of Ro-Man, pages 55–60.

Goodman, P.S., Ravlin, E., & Schminke, M. (1987) Understanding groups in organizations. In L. L. Cummings & B.M. Staw (Eds.), Research in organizational behavior (vol. 9, pp. 121-173). Greenwich, CT:JAI Press.

Grosz,B.& Sidner, C. (1990) Plans for discourse. In P. Cohen et al., eds., Intentions in Communication.MIT Press.

Hancock, P., Billings, D., Schaefer, K., Chen, J., De Visser, E. & Parasuraman, R. (2011) A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. Human Factors,53(5)517-527.

Holland, O., Diamond, A., Marques, H. G., Mitra, B., & Devereux, D. (2013). Real and apparent biological inspiration in cognitive architectures. Biologically Inspired Cognitive Architectures, 105–116.

Johnson-Laird, P.N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (2006) *How We Reason*. Oxford University Press

Kaupp, T. & Makarenko, A. (2008) Measuring human-robot team effectiveness to determine an appropriate autonomy level. IEEE International Conference on Robotics and Automation, May 2008, pp.2146-2151

Kim, K. G. Cha, Y. S. Park J. M., Lee, J. Y. & You, B. J. (2011) Providing services using network-based humanoids in a home environment. IEEE Trans. Consumer Electronics , vol. 57, no. 4, pp. 1628-1636.

Kirkman, B.L. & Rosen, B. (1999) Beyond self-management: antecedents and consequences of team empowerment. Academy of Management Journal, 42(1), 58-74.

Levesque, H. J., Cohen, P. R., & Nunes, J. (1990) On acting together. In Proceedings of the National Conference on Artificial Intelligence. Menlo Park, Calif.: AAAI press.

Linegang, M., Haimson, C., MacMillan, J. & Freeman, J. (2003) Human control in mixed-initiative systems: lessons from the MICA-SHARC program. IEEE International Conference on Systems, Man and Cybernetics, 2003, Volume 1, pp.436-441.

Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Salas, E., & Cannon-Bowers, J.A. (2000) The influence of shared mental models on team process and performance. Journal of Applied Psychology, 85(2), 273-283.

Mullen, B. & Cooper,C. (1994) The relation between group cohesiveness and performance: An integration. United States Army Research Institute for the Behavioral and Social Sciences. ARI Research Note - 95-31.

Murphy, R. R. & Rogers, E. (2001). Human-Robot Interaction. Available at http://www.aic.nrl.navy.mil/hri/nsfdarpa/

Nevatia, Y., Stoyanov, T., Rathnam, R., Pfingsthorn, M., Markov, S., Ambrus, R. & Birk, A. (2008) Augmented autonomy: Improving human-robot team performance in Urban search and rescue. IEEE/RSJ International Conference on Intelligent Robots and Systems, Sept. 2008, pp.2103-2108

O'Brien, D. (2009) Human reasoning includes a mental logic. *Behavioral Brain Science.* 32, 96–97.

Parasuraman et al (2000)

Prakash, A., Beer, J. M., Deyle, T., Smarr, C.-A., Chen, T. L., Mitzner, T. L., Kemp, C. C., & Rogers, W. A. (2013) Older adults' medication management in the home: How can robots help? 8th ACM/IEEE International Conference on Human-Robot Interaction, March 2013, pp.283-290.

C. Rich, C. Sidner and N. Lesh (2001) "Collagen: Applying Collaborative Discourse Theory to Human-Computer Interaction." *AI Magazine*. Winter 2001. pp 15-25.

Richards, D. & Stedmon, A. (2015) To delegate or not to delegate: A review of control frameworks for autonomous cars. Applied Ergonomics, 53, 383-388.

Richards, D., Stedmon, A., Shaikh, S. & Davies, D. (2014) Responding to disaster using autonomous systems. The Ergonomist, No.534, December 2014, pp.8-9.

Rips, L.J. (1994). The psychology of proof. Deductive reasoning in human thinking. Cambridge, Massachutsetts, MIT Press, Bradford

Salas, E., Cooke, N. J. & Gorman, J.C. (2010) The science of team performance: Progress and the need for more... Human Factors, 52, 344-346.

Salas, E., Stagl, K. C., Burke, C. S., & Goodwin, G. F. (2007). Fostering team effectiveness in organizations: Toward an integrative theoretical framework of team performance. In R. A. Dienstbier, J. W. Shuart, W. Spaulding, & J. Poland (Eds.), Modeling complex systems: Motivation, cognition and social processes: Nebraska Symposium on Motivation (Vol. 51, pp. 185–243). Lincoln: University of Nebraska Press.

Sarter, N.B. & Woods, D.D. (1997) Team play with a powerful and independent agent: Operational experiences and automation surprises on the Airbus A-320. Human Factors, 39, 553-569.

Schermerhorn P. & Scheutz, M. (2011) Disentangling the effects of robot affect, embodiment, and autonomy on human team members in a mixed-initiative task. In Proceedings of the 2011 International Conference on Advances in Computer-Human Interactions, Gosier, Guadeloupe, France.

Scholtz, J. (2003) Theory and Evaluation of Human Robot Interactions. Hawaii International Conference on System Science 36 (HICSS 36). Hawaii, January 2003.

Schultz, A. & Parker, L.E. (Eds). (2002). Multi-Robot Systems: From Swarms to Intelligent Automata—Proceedings from the 2002 NRL Workshop on Multi-Robot Systems. Boston: Kluwer Academic Publishers.

Sheridan, Thomas B. 1992, Telerobotics, Automation, and Human Supervisory Control. Cambridge, MA: MIT Press.

Sheridan, T. B. (2011) Adaptive automation, level of automation, allocation authority, supervisory control, and adaptive control: Distinctions and modes of adaptation. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 41(4), 662-667.

Sheridan, T. B., & Verplank, W. L. (1978). Human and computer control of undersea teleoperators. Man-Machine Systems Laboratory Report. Cambridge, MA: MIT.

Shiomi, M., Kanda, T., Ishiguro, H. & Hagita, N. (2006) Interactive humanoid robots for a science museum. In Proceedings of 1st ACM/IEEE International Conference of Human–Robot Interaction, 2006, pp. 305–312.

Singer, S.M. & Akin, D.L. (2010) Scheduling robot task performance for a cooperative human and robotic team. Acta Astronautica, 66, 102-116.

Thiel, S., Habe, D. & Block, M. (2009) Co-operative robot teams in a hospital environment. Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on , vol.2, no., pp.843-847, 20-22.

Thomas, K.W. (1992) Conflict and conflict management: Reflections and update. Journal of Organizational Behavior, 13(3), 265-274.

Tiferes, J., Hussein, A.A., Bisantz, A., Kozlowski, J.D., Sharif, M.A., Winder, N.M., Ahmad, N.A., Allers, J., Cavuto, L. & Guru, K.A. (2016) The loud surgeon behind the console: Understanding team activities during robot-assisted surgery. Journal of Surgical Education, 73(3), 504-512.

Trafton, J. G., Hiatt, L. M., Harrison, A. M., Tamborello, P., Khemlani, S. S., & Schultz, A. C. (2013). Act-r/e: An embodied cognitive architecture for human–robot interaction. Journal of Human–Robot Interaction, 2, 30–54.

Webber, S.S. (2002) Leadership and trust facilitating cross-functional team success. Journal of Management Development, 21(3), 201-214.

White, A. & Richards, D. (2006) Building operator trust in semi-autonomous UCAV systems. 1st 'Moving Autonomy Forward' Conference, 21-22 June 2006, UK: Grantham.

Acknowledgements

Thanks go to Davide Secchi, Associate Professor of Organisational Cognition at the University of Southern Denmark, for inviting this paper after parts of this work was presented at the AISB Workshop Agent-Based Models of Bounded Rationality at University of Southern Denmark, 7-8th May 2015. Thanks also to Professor Petru Curseu for his helpful comments on an earlier draft of this paper.