Designing for Mixed-Initiative Interactions between Human and Autonomous Systems in Complex Environments

Michael J. Barnes
Human Research & Engineering Directorate
U.S. Army Research Laboratory
Ft. Huachuca, AZ, USA
michael.j.barnes.civ@mail.mil

Jessie Y. C. Chen
Human Research & Engineering Directorate
U.S. Army Research Laboratory
Orlando, FL, USA
yun-sheng.c.chen.civ@mail.mil

Florian Jentsch
Psychology Department and Institute for Simulation & Training
University of Central Florida
Orlando, FL, USA
Florian.Jentsch@ucf.edu

***Abstract*—The purpose of the paper is to discuss human centered design implications for shared decision making between humans and autonomous systems in complex environments. Design implications are generated based on empirical results from two research paradigms. In the first paradigm, an intelligent agent (RoboLeader) supervised multiple subordinate systems and was in turn supervised by the human operator. The RoboLeader research varied number of subordinate units, task difficulty, agent reliability, type of agent errors, and partial autonomy. The second paradigm involved human interaction with partially and fully autonomous systems. Design implications from both paradigms are evaluated—relating to multitasking, adaptive systems, false alarms, individual differences, operator trust, and allocation of human and agent tasks for partial autonomy. We conclude that mixed-initiative decision sharing depends on designing interfaces that support human-agent transparency.**

Keywords-autonomous agents; human-centered design; mixed-initiative decision making

I. INTRODUCTION

The dictionary definition of autonomy is "not subject to control from outside; independent, existing and functioning as an independent organism" [1]. Although there are many definitions of autonomous systems [3], we define the concept in a more circumscribed manner indicating an intelligent agent architecture that is able to conduct relatively complex missions, is able to adapt to its environment in predictable ways, and is able to achieve programmed objectives [3]. A car that is able to navigate on the freeway and find the off-ramp and arrive at its final destination would meet our definition but this does not imply that the agent makes important decisions without permission or is not supervised by a human. Thus we are assuming an important human role – the system will require some level of supervision but not necessarily continual monitoring. Further, in complex environments, particularly military ones, a single operator may be responsible for multiple unmanned systems thus expanding the operator's role to even more complex situations. This paper will discuss research at the U.S. Army Research Laboratory (ARL) and funded projects that focused on human issues related to autonomy [2]-[3]. In particular, we will discuss models of adaptation and argue that mixed-initiative architectures are an ideal compromise between adaptive and adaptable approaches. We will discuss RoboLeader as exemplar of a mixed-initiative architecture for the control of multiple systems and discuss the corpus of research indicating important human-autonomy interface design issues [15]-[17]. The research includes span of control, agent reliability, and levels of autonomy (LOA), task loadings and the effects of individual differences on human-agent teaming. We will discuss collaborative work at the University of Central Florida (UCF) investigating various aspects of autonomy showing advantages and disadvantages of different approaches [4]-[8]. Further, we will review results of other ARL funded research directly related to developing user interfaces (UI) for autonomous systems with emphasis on their implications for design of human-agent collaboration [3].

A. Models and Architectures of Human/Agent Adaption

The problem we address is how to adapt to complex environments with a many-to-one ratio of systems to humans [9]. Autonomy is necessary but not sufficient for such systems. As the number of systems to be supervised increases beyond the human's short term memory limitations, the operator loses the ability to keep track of the independent entities. The number of robots that can be monitored is particularly limited in adverse situations with an operator being able to supervise as few as three systems [9], [10]. The human's attentional focus is limited by the number of independent entities that are being monitored even if they are not being directly controlled; thus autonomy as described above does not ensure effective supervision. This is particularly true in complex, possibly volatile, environments such as driving on a Los Angeles freeway or being engaged in combat. Also, if multiple heterogeneous systems are involved, the level of complexity increases as function of the number of interacting entities and the nature of the interfaces [11]. In a literature review [2], we have identified three possible types of interaction between humans and agents that reduce workload: adaptive [12], adaptable [13] and mixed-initiative models [14]. Adaptive assumes that triggers in the environment causes the agent to respond without operator intervention; adaptable assumes that a trigger causes the agent to suggest a pre-programmed action; and mixed-initiative assumes a collaborative decision process. Having both the human and agent as decision-makers captures the advantage of adaptive and adaptable processes without sacrificing the human's decision authority [3], [14]. The mixed-initiative architecture allows

978-1-4799-8697-2/15 $31.00 © 2015 IEEE
DOI 10.1109/SMC.2015.246

humans to oversee agent decisions while allowing agents to respond to emergencies and conduct lower-level tasks such as convoy spacing with minimal oversight. Because of our interest in military systems, we stress the importance of having the human in the decision-loop in order to take advantage of the human's greater flexibility and understanding of political and ethical considerations [2]. To solve the n-system problem, we have proposed a concept of a supervisory agent: RoboLeader [15] as an exemplar of a hierarchical multi-agent architecture that has mixed-initiative relationship with the human operator [2], [16]-[17].

II. RoboLeader

RoboLeader was designed to be an software emulation of a supervisory agent who has purview of 4-8 aerial or ground systems that operate autonomously unless there is an unexpected event such as blocked road or a no/go area as indicated by a recent intelligence update [15]. In a surveillance task, an individual robot must not only react to events such as the blocked road but must chose solutions that are optimal for the set of autonomous vehicles to cover the assigned area. The literature suggest a number of algorithmic solutions such as market-based algorithms (see [2], [28] for reviews); however, our experimental focus was on the interaction between the agent supervisor and the human operator—not on algorithmic development. Figure 1 shows the user interface for one of the simulations. The text box and symbols on the map indicates the suggested solution generated by RoboLeader to circumvent obstacles. The graphical UI provides the information the operator needed to oversee RoboLeader's solution by overtly agreeing or disagreeing with RoboLeader's suggested course of action. The initial experiments indicated that RoboLeader reduced decision time by 13% for agent supervision of 4-8 robots conducting surveillance missions when the agent's suggested route changes were 100% accurate.

Figure 1. RoboLeader user interface showing chat window, targeting videos, gauge monitoring, and map [16].

In the second experiment, Chen and Barnes [16] varied agent reliability level, type of errors (False alarm prone [FAP] or miss prone [MP]) and target density. To make the simulation realistic, they required participants to detect targets as the robots traversed the surveillance routes, change routes when obstacles prevented continuation on the pre-programmed routes and answer situation awareness (SA) probes throughout the experimental sessions. Symbols on the map gave the operators indicators of obstacles requiring a route change, and the robot's sensor view of its current route showed possible targets on the robot's video window (Figure 1). Figure 2 indicates that FAP errors were less deleterious for target acquisition than MP errors. This seems to contradict previous literature indicating either a trade-off between MP or FAP effects or generally negative effects of FAP type of errors [18]; whereas Chen and Barnes [16] found FAP generally superior to MP conditions in the automated task (route changes) as well as the concurrent target detection task. The difference in the paradigms seem to be the salience of FAP icons which were easily checked on the user interface whereas the MP conditions required continually checking. The SA data were the exception, the MP resulted in better performance suggesting that participant's SA benefited from the continual search of the UI required when the agent's miss rate was high (Figure 2). The implication is that effects due to error types are influenced by display design, FAP errors that are easily checked even for low reliability conditions (60%) are less harmful than MP agents that require constant oversight. The target detection performance and operator SA data from both RoboLeader experiments are contrasted in Figure 2.

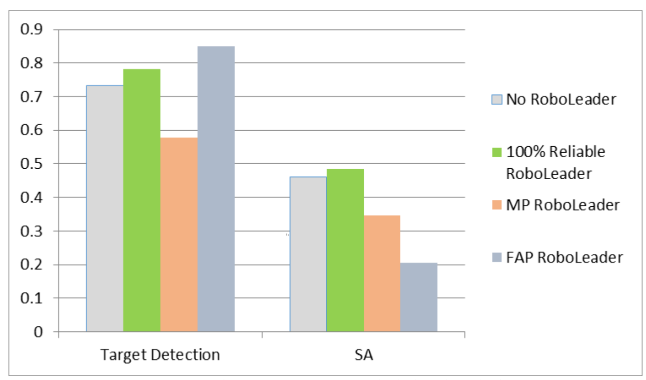

Figure 2. Target Detected as a function of reliability level, density and False Alarm (FAP) and Miss Prone (MP) reliability functions.

Additional RoboLeader experiments investigated partial autonomy (human helped with final decision) compared to full autonomy (agent decision) compared to manual (human only). Chen and Barnes [17] examined a more difficult task involving the robots entrapping an escaping vehicle much like police vehicles cornering a vehicle suspected of criminal activity (Figure 3). The RoboLeader agent suggested a partial solution vs. an autonomous solution. The agent conditions were superior to the manual condition whereas partial autonomy results suggested that there were benefits to having humans involved in the entrapment task (humans helped position the robots in the final entrapment step). However, research by Pettit et al. [19] at

Ft. Benning showed the opposite effect: splitting tasks between the human and autonomous control of a small robot resulted in performance decrements compared to full autonomy. This was most likely due to a mismatch between the autonomous component and the task assigned to the human in the partial condition. Pettit et al.'s experiment required participants to take control of the robot to move around an obstacle but navigation was autonomous. Their participants reported that this was particularly difficult for them to do compared to full autonomy. In Chen and Barnes' paradigm [17], the operator made a position decision on where the robots would end up to entrap the rouge vehicle but the operator was not required to take control of the robots.

Figure 3. RoboLeader user interface [17].

In a recent study, Wright et al. [20] investigated partial autonomy using a different RoboLeader paradigm (Figure 4). The operator assumed the role of a convoy commander who performed under different RoboLeader autonomy conditions. Participants conducted a surveillance task with the participant having a 360^0 view of the urban terrain fore and aft of the lead vehicle displaying multiple possible targets. Three levels of RoboLeader assistance were examined – (L1) none (operator manually completing the convoy management tasks), (L2) vehicle separation, and (L3) vehicle separation and route changes. The manual condition had the poorest performance whereas L3 and L2 were similar except for an increase in false alarms for level 2. Workload as assessed by the NASA-Task Load Index [21] indicated that not only did both autonomy conditions reduce workload but that the operators reported more overall workload in L2 than L3, suggesting partial autonomy had mixed results in this experiment.

Wright et al. [20] also used eye tracking data that had been shown to be correlated with workload in previous research to parse out effects of individual differences. The general trend was an increase in fixation count (FC) and a decrease in pupil diameter (PD–a measure of global cognitive engagement) for the autonomy conditions compared to the manual condition. The most likely interpretation is that autonomy allowed for a wider search pattern (FC) whereas the PD indicator implies autonomy reduced participant's workload. However, individual differences in perceived attentional control (AC) for the PD and FC measures suggest a more complicated picture. Those with lower AC had significantly smaller PD (workload) that were unaffected by autonomy level. The higher AC participants had larger PD that were reduced for the autonomous conditions. The FC showed the opposite trend: higher FC for the lower AC participants compared to the higher AC participants. This indicates that individual differences in AC mitigated both how they responded to autonomy (no PD differences for low AC) and search patterns (higher FC for low AC), suggesting different workload strategies between the two AC sub-groups.

Figure 4. RoboLeader user interface [20].

More generally, both in the literature and in RoboLeader findings, individual differences are an essential element of human agent interaction [2]. In a number of the RoboLeader experiments, gamers and those with superior spatial abilities were able to scan the UI more efficiently, and gamers in particular were able to maintain information in short term visual memory better than infrequent gamers [16], [17]. Operators with high AC scores were able to attend to secondary tasks more efficiently suggesting that they were able to effectively split their attention between the primary tasks (supervision of RoboLeader and target detection) and the other tasks (gauge monitoring, SA and communications) during multitasking. Taken together, the results suggest that training and decision support need to be designed for individual differences and not for average proficiency.

III. LOA AND PERFORMANCE

The University of Central Florida (UCF) in collaboration with ARL investigated effects of varying levels of autonomy (LOA) and different types of autonomy on human robot interaction during simulated combat mission [4]-[8]. In their first series of experiments, they used a $1/35^{th}$ scale mid-eastern urban area, in which the scaled robots traversed the model's streets populated by vehicles and similarly scaled human figures (Figure 5). The initial studies looked at robot-to-robot automated interactions compared to the human controlling both robots [5]. In the first experiment, the robots had to change routes either by communicating with robot peers or under human control. For the easy tasks, when communications were less than 100% reliable, manual control was superior to robot-to-robot control, but when various impediments such as no/go zones were introduced, then robot-to-robot automated route finding improved performance. Thus the efficacy of introducing automation depended on task difficulty as well as task type. The designer must find a sweet spot based on task difficulty between manual control and some level of autonomy. One possibility would be to introduce an adaptive (or adaptable) algorithm triggered by workload increases as a means of eliciting automated aiding as a function of task difficulty [2], [13].

Figure 5. Scaled facility for military operations in urban terrain.

Another study [6] decomposed the robot's task into three components (a) object detection, (b) object identification, and (c) decision making based on military rules of engagement (ROE). The scenario consisted of the robot operating under different LOAs while navigating on a road, on which various obstacles were encountered. After detecting and identifying each obstacle, the participant had to decide whether to continue on the route based on pre-briefed ROEs. There were three experimental conditions: manual, collaborative and autonomous. The fully autonomous condition was best for detection–almost doubling the ability of the robot's to detect and stop safely when encountering an obstacle compared to manual detections. Although the robots speed was relatively slow (~ 5 mph), the ability of the operator to spot obstacles and react in an urban environment was poor (37%) implying that safe obstacle avoidance needs to be automated. For the higher cognitive level tasks (b and c above) performance was better for the collaborative tasks. In particular, the "Perception by Proxy" paradigm for object identification resulted in effective human collaboration with autonomy. In this paradigm, the autonomy detected the object and participants used their perceptual knowledge to identify the object and understand its military significance.

In the final experiment reviewed, Jentsch et al. [8] varied both reliability level (60% and 90%) and the type of human autonomy interaction (full autonomy, management by consent [MC], and management by exception [ME]). In MC, the operator had to signal agreement with agent's decision, whereas ME only required the operator not to override the agent's decision. Averaged over both reliability conditions, ME resulted in the best performance. This was because the operators did not second guess the autonomy suggestion for high reliability conditions but was able to override the autonomy when it was obviously wrong especially in the low reliability conditions. Apparently participants in the MC condition over-trusted their ability compared to the autonomous agent even in conditions when the agent was highly effective. This interpretation was buttressed by the subjective trust data indicating that although overall, the MC condition resulted in the worst performance, participants trusted the MC system the most. In this condition, they had the most control over whether to use the autonomy because the simulated robot was halted until the operator made a decision.

The design implications from these studies imply that partial autonomy must be designed with the relative strengths and weaknesses of the agents and human operators. In particular, detecting an object on a road is an ideal task for autonomous algorithms of even moderate effectiveness. This is a task that must be done rapidly with little margin for error. However, both identification and classification and decision making are best accomplished as a collaborative task taking advantage of the wider range of human knowledge. A counterintuitive and important finding is that subjective trust can mislead an operator into over trusting either his own or the autonomous systems capabilities. Finally, especially under low workload conditions, there are many situations where manual control may be preferred to autonomy. Adaptive systems that take into account the state of the operator or the environment may be an efficient way to share autonomy with its human operator especially with the added agency of a mixed-initiative system [2].

IV. Conclusions and Future Research

We covered important design implication for mixed-initiative systems including multitasking, interface issues, type of automation errors, partial autonomy, trust and individual differences in supervisory control. Our discussion indicated both advantages [2], [16], [17] and disadvantages of shared decision paradigms [7], [8]. For example, participants found controlling a moving vehicle difficult compared to autonomous conditions but they were better at identifying the significance of an object or making more abstract decisions [4], [6], [17]. To summarize, architectures that split decision making between autonomous agents and human are generally (but not always) more effective than an architecture that depends on one or the other. There needs to multiple trade-offs among decision precision, task allocations, human oversight, and manageable operator workload. Such a decision architecture is complicated and its success is predicated on not only the issues discussed in the paper but also on further research as discussed in the following paragraphs.

We conclude by arguing that a number of important issues need to be understood before mixed-initiative systems are realistic in complex environments. The system UI must provide transparency by indicating the agent's intentions and its reasoning [23]. Especially in novel situations, an opaque agent in an ever-changing combat environment will not be trusted. Assuming either that the agent is always correct or ignoring the agent when it is correct can lead to catastrophic results [24]. Chen and colleagues currently are investigating transparency concepts that increase the operator's awareness into what the agent intends, its logic, and the predicted results during a complex base defense scenario [25].

The other important research issue that ARL and UCF researchers are working on is bi-directional communication with an agent peer. The agent must know what the human intends and the agent needs to be able to query its operator concerning possible changes or ambiguities in the environment [2]. Elliott and her colleagues at Ft. Benning are working on naturalistic interfaces for dismounted soldiers using gesture or tactile communication to interact with robots. The proposed equipment solutions are light-weight and do not depend on a hand-held visual displays, permitting soldiers to interact with robots without losing SA [26], [27]. Elliott is also working with UCF researchers evaluating bi-directional voice and gesture communications between a small robot and soldiers conducting a joint mission. Future efforts will need to scale up bi-directional communications to more complex situations involving multiple autonomous systems [2].

ACKNOWLEDGMENT

This research was supported by the U.S. Army Research Laboratory's human-robot interaction program led by Susan Hill. We wish to thank all the researchers who contributed to the research findings discussed in this paper especially Daniel Barber, Linda Elliott, Thomas Fincannon, and Julia Wright.

REFERENCES

[1] Dictionary.com, "Autonomous," in Dictionary.com Unabridged.

[2] J. Y. C. Chen and M. J. Barnes, "Human-agent teaming for multi-robot control: A review of human factors issues." IEEE Trans. Human-Machine Syst., vol. 22, Feb. 2014, pp. 13-29.

[3] M. J. Barnes, J. Y. C. Chen, F. Jentsch, and E. S. Redden, "Designing effective soldier-robot teams in complex environments: Training, interfaces, and individual differences," in Eng. Psychology and Cognitive Ergonom., Lecture Notes in Comput. Sci., vol. 6781, D. Harris, Ed. Berlin: Springer-Verlag, 2011, pp. 484–493.

[4] T. Fincannon, A. W. Evans III, F. Jentsch, and J. Keebler, "Dimensions of spatial ability and their influence on performance with unmanned syst.," in Human Factors in Defense: Human Factors in Combat Identification, D. H. Andrews, R. P. Herz, and M. B. Wolf, Eds. Aldershot, UK: Ashgate, 2011, pp. 67–81.

[5] T. Fincannon, J. R. Keebler, F. Jentsch, E. Phillips, and A. W. Evans, "Team size, team role, commun. modality, and team coordination in the distributed operation of multiple heterogeneous unmanned vehicles" J. Cogn. Eng. Decision Making, vol. 5, March 2011, pp. 106–131.

[6] F. Jentsch and T. Fincannon, "Multiple robot-multiple operator control and teamwork," Proc. Inform. Technology Conf. American Instit. for Aeronautics and Astronautics, AIAA, 2012.

[7] B. Sellers, T. Fincannon, and F. Jentsch. "The effects of autonomy and cognitive abilities on workload and supervisory control of unmanned systems," Proc. Human Factors and Ergonomics Society 56th Annual Meeting, HFES, Oct. 2012, pp. 1039-1043.

[8] F. Jentsch, T. Fincannon, B. Sellers, J. Keebler, S. Ososky, E. Phillips, and D. Schuster, "Safe operations of autonomy, teaming and workload operator trust and performance," Univ. of Central Florida, Orlando, FL, Tech. Rep. TO-132-12/11, 2011.

[9] M. Lewis and J. Wang, "Coordination and automation for controlling robot teams," in Human Robot Interactions in Future Military Operations, M. J. Barnes, F. Jentsch, Eds. Farnham, Surrey, UK: Ashgate, 2010, pp. 398-418.

[10] G. Miller, "The magical number 7 plus or minus two: Some limits on our capacity for processing information," Psychological Review, vol. 63, 1956. pp. 81–97.

[11] M. Hou, H. Zhou, M. Zhou, and G. Arrabito, "Optimizing operator agent interaction in intelligent adaptive interface design: A conceptual framework," IEEE Trans. Syst., Man, Cybern. C, App. Rev., vol. 41, Mar. 2011, pp. 161–178.

[12] R. Parasuraman, M. Barnes, and K. Cosenzo, "Adaptive automation for human-robot teaming in future command and control systems," Int. C2 Journal, vol. 1, 2011, pp 43–68.

[13] C. Miller and R. Parasuraman, "Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control," Human Factors, vol. 49, Feb. 2007, pp. 7–63.

[14] M. Goodrich, "On maximizing fanout: Towards controlling multiple robot teams," in Human Robot Interactions in Future Military Operations, M. J. Barnes, F. Jentsch, Eds. Farnham, Surrey, UK: Ashgate, 2010, pp. 368–393.

[15] Z. Snyder, Qu, J. Y. C. Chen, and M. J. Barnes, "RoboLeader for reconnaissance by a team of robotic vehicles," Proc. Int. Symp. Collab. Tech. Syst. (CTS 10), 2010, pp. 522–530.

[16] J. Y. C. Chen and M. J. Barnes, "Supervisory control of multiple robots: Effects of imperfect automation and individual differences," Human Factors, vol. 54, Apr. 2012, pp. 157–174.

[17] J. Y. C. Chen and M. J. Barnes, "Supervisory control of multiple robots in dynamic tasking environments," Ergonom., vol. 59, Sept. 2012, pp. 1043–1058.

[18] S. R. Dixon, C. D. Wickens, and J. M. McCarly, "On the independence of reliance and compliance: Are false alarms worse than misses," Human Factors, vol. 49, Aug. 2007, pp. 564-572.

[19] R. A. Pettit, E. S. Redden, C. B. Carstens, and D. Hooper, Scalability of robotic controllers: Effects of progressive autonomy on intell., surveillance, and reconnaissance robotic tasks (Technical Report: ARL-TR-6182). Aberdeen Proving Ground MD: US Army Research Laboratory, 2012.

[20] J. L. Wright, J. Y. C. Chen, and S. A. Quinn, and M.J. Barnes, The effects of level of autonomy on human-agent teaming for multiple-robot control and local security maintenance (Technical Report: ARL-TR-6724). Aberdeen Proving Ground MD: US Army Research Laboratory, 2013.

[21] S. Hart and J. Staveland, "Development of NASA TLX: Results of empirical and theoretical research," in Human Mental Workload, P. Hancock and N. Meshkati, Eds. Amsterdam: Elsevier, 1998, pp. 139-183.

[22] W. S. Peavler, "Pupil size,inforamtion overload, and performance differences" Psychophysiolgy, vol. 11, 1974, pp.559-566.

[23] J. D. Lee and K. A. See, "Trust in technology: Designing for appropriate reliance," Human Factors, vol. 46, Mar. 2004, pp. 50–80.

[24] R. Parasuraman and V. Riley, "Humans and automation: use, misuse, disuse, abuse," Human Factors, vol. 39, Jun. 1997, pp. 230–253.

[25] J. Y. C. Chen, K. Procci, M. Boyce, J. Wright, A. Garcia, and M. Barnes, Situation awareness-based Agent Transparency (Technical Report: ARL-TR-6905). Aberdeen Proving Ground MD: US Army Research Laboratory, 2014.

[26] L. R. Elliott, C. Jensen, E. S. Redden, and E. Pettit, Robotic telepresence: Perception, performance and user experience (Technical Report: ARL-TR-5928). Aberdeen Proving Ground MD: US Army Research Laboratory, 2012.

[27] L. R. Elliot, B. Mortimer, G. Hartnett-Pomranky, G. Zets, and G. Mortz, "Augmenting soldier situation awareness and navigation through tactile cueing," Proc. International Conference on Human-Computer Interaction, in press.

[28] M. Lewis, "Human interaction with multiple remote robots," Rev. Human Factors Ergonom., vol. 9, 2013, pp. 131–173.