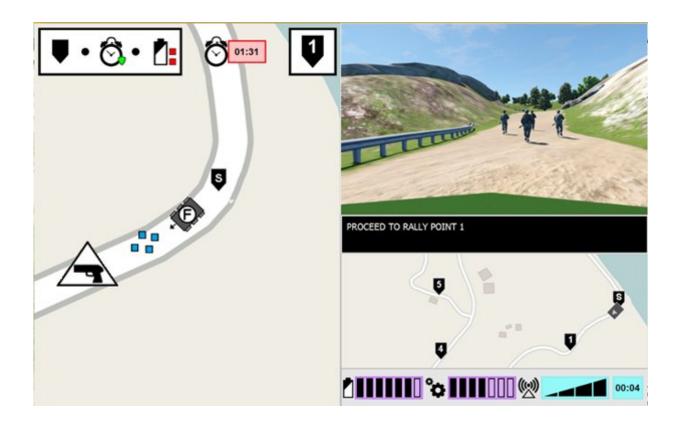# Army scientists improve human-agent teaming by making AI agents more transparent

January 11 2018

The Autonomous Squad Member is a small ground robot that interacts with and communicates with an infantry squad. As part of the overall ASM program, Chen and colleagues developed transparency visualization concepts, which they used to investigate the effects of agent transparency levels on operator performance. Informed by the SAT model, the ASM's user interface features an '"at a glance" transparency module' where user-tested iconographic representations of the agent's plans, motivator, and projected outcomes are used

to promote transparent interaction with the agent. Credit: Dr. Jessie Y. Chen

U.S. Army Research Laboratory scientists developed ways to improve collaboration between humans and artificially intelligent agents in two projects recently completed for the Autonomy Research Pilot Initiative supported by the Office of Secretary of Defense. They did so by enhancing the agent transparency, which refers to a robot, unmanned vehicle, or software agent's ability to convey to humans its intent, performance, future plans, and reasoning process.

"As machine agents become more sophisticated and independent, it is critical for their human counterparts to understand their intent, behaviors, reasoning process behind those behaviors, and expected outcomes so the humans can properly calibrate their trust in the systems and make appropriate decisions," explained ARL's Dr. Jessie Chen, senior research psychologist.

The U.S. Defense Science Board, in a 2016 report, identified six barriers to human trust in autonomous systems, with 'low observability, predictability, directability and auditability' as well as 'low mutual understanding of common goals' being among the key issues.

In order to address these issues, Chen and her colleagues developed the Situation awareness-based Agent Transparency, or SAT, model and measured its effectiveness on human-agent team performance in a series of human factors studies supported by the ARPI. The SAT model deals with the information requirements from an agent to its human collaborator in order for the human to obtain effective situation awareness of the agent in its tasking environment. At the first SAT level, the agent provides the operator with the basic information about its current state and goals, intentions, and plans. At the second level, the

agent reveals its reasoning process as well as the constraints/affordances that the agent considers when planning its actions. At the third SAT level, the agent provides the operator with information regarding its projection of future states, predicted consequences, likelihood of success/failure, and any uncertainty associated with the aforementioned projections.

In one of the ARPI projects, IMPACT, a research program on human-agent teaming for management of multiple heterogeneous unmanned vehicles, ARL's experimental effort focused on examining the effects of levels of agent transparency, based on the SAT model, on human operators' decision making during military scenarios. The results of a series of human factors experiments collectively suggest that transparency on the part of the agent benefits the human's decision making and thus the overall human-agent team performance. More specifically, researchers said the human's trust in the agent was significantly better calibrated—accepting the agent's plan when it is correct and rejecting it when it is incorrect— when the agent had a higher level of transparency.

The other project related to agent transparency that Chen and her colleagues performed under the ARPI was Autonomous Squad Member, on which ARL collaborated with Naval Research Laboratory scientists. The ASM is a small ground robot that interacts with and communicates with an infantry squad. As part of the overall ASM program, Chen's group developed transparency visualization concepts, which they used to investigate the effects of agent transparency levels on operator performance. Informed by the SAT model, the ASM's user interface features an at a glance transparency module where user-tested iconographic representations of the agent's plans, motivator, and projected outcomes are used to promote transparent interaction with the agent. A series of human factors studies on the ASM's user interface have investigated the effects of agent transparency on the human

teammate's situation awareness, trust in the ASM, and workload. The results, consistent with the IMPACT project's findings, demonstrated the positive effects of agent transparency on the human's task performance without increase of perceived workload. The research participants also reported that they felt the ASM as more trustworthy, intelligent, and human-like when it conveyed greater levels of transparency.

Chen and her colleagues are currently expanding the SAT model into bidirectional transparency between the human and the agent.

"Bidirectional transparency, although conceptually straightforward—human and agent being mutually transparent about their reasoning process—can be quite challenging to implement in real time. However, transparency on the part of the human should support the agent's planning and performance—just as agent transparency can support the human's situation awareness and task performance, which we have demonstrated in our studies," Chen hypothesized.

The challenge is to design the user interfaces, which can include visual, auditory, and other modalities, that can support bidirectional transparency dynamically, in real time, while not overwhelming the human with too much information and burden.

Provided by U.S. Army Research Laboratory

Citation: Army scientists improve human-agent teaming by making AI agents more transparent