Measuring Complacency in Humans Interacting with Autonomous Agents in a Multi-Agent System

Sebastian S. Rodriguez[a], Jacqueline Chen[a], Harsh Deep[a], Jaewook (Jae) Lee[a], Derrik E. Asher[b], and Erin Zaroukian[b]

[a]Department of Computer Science, University of Illinois at Urbana-Champaign, 201 N Goodwin Ave., Urbana, IL USA 61801
[b]CCDC Army Research Laboratory, 2800 Powder Mill Rd., Adelphi, MD USA 20783

ABSTRACT

With advances in machine learning, autonomous agents are increasingly able to navigate uncertain operational environments, as is the case within the multi-domain operations (MDO) paradigm. When teaming with humans, autonomous agents may flexibly switch between passive bystander and active executor depending on the task requirements and the actions being taken by partners (whether human or agent). In many tasks, it is possible that a well-trained agent's performance will exceed that of a human, in part because the agent's performance is less likely to degrade over time (e.g., due to fatigue). This potential difference in performance might lead to complacency, which is a state defined by over-trust in automated systems. This paper investigates the effects of complacency in human-agent teams, where agents and humans have the same capabilities in a simulated version of the predator-prey pursuit task. We compare subjective measures of the human's predisposition to complacency and trust using various scales, and we validate their beliefs by quantifying complacency through various metrics associated with the actions taken during the task with trained agents of varying reliability levels. By evaluating the effect of complacency on performance, we can attribute a degree of variation in human performance in this task to complacency. We can then account for an individual human's complacency measure to customize their agent teammates and human-in-the-loop requirements (either to minimize or compensate for the human's complacency) to optimize team performance.

Keywords: human-autonomy teaming, predator-prey pursuit, automation bias, automation complacency, group dynamics, adaptive agents

1. INTRODUCTION

Warfighters are taught from indoctrination that "there's no I in team." Teams are organizations that employ dynamic and adaptive behavior between individuals in order to achieve a common goal.[1] For warfighters, good team performance can be the difference between mission success and failure. With advances in computational technologies, machine learning has allowed us to develop autonomous agents: non-living entities which have a capacity to be intelligent and make their own decisions. Because of this, the fundamental structure of teaming has changed – teams can now be comprised of a combination of human members and autonomous agents.[2] With automation and autonomy being ubiquitous in the 21st century, human-autonomy teams already exist not only for specialized workers, but also in the ordinary person's daily life (e.g., autonomous vehicles, content recommendations, algorithmic decision-making systems), resulting in a growing need to study their interactions.[3] In turn, our service members increasingly rely on automation to complete missions, requiring effective computation agents to be human-aware in order to behave like teammates rather than tools.[4]

With autonomy growing adaptive and cognisant of its operating environment, it can become complex and unpredictable. This presents a challenge, as autonomy is currently unable to effectively communicate with its team members without specialized and task-specific protocols,[5] which falls outside of Salas et al.'s 3C model of teamwork.[6] Teamwork hinges upon team members being able to coordinate (e.g., adopt strategies based on observations of partners' behaviors), cooperate (e.g., optimize cohesion based on shared objectives), and

Corresponding author: Sebastian S. Rodriguez (srodri44@illinois.edu)

communicate (e.g., provide feedback), and albeit the lack of communication does not imply the degradation of teamwork, it may not be as effective as a team comprised entirely of humans.

Research has taken advantage of autonomy and computation in order to provide capabilities that match or surpass their human teammates. As autonomous teammates evolve from being passive bystanders (e.g., recommender and decision support systems) to active executors of a task, it is possible that performance of the agents grow to exceed the human's, in part because the agent's performance is less likely to degrade over time.[7] This potential difference might lead to complacency, which is a state defined by over-trust in automated systems.[8] A human might over-trust the capabilities of their autonomous teammates, resulting in reduced situation awareness.[9] Following the paradigm of social loafing (an under-researched area of Human-Autonomy Teaming), is the effect of a human's potential over-reliance toward autonomous teammates, and the complacency that results from over-trust due to the system's perceived reliability. Even with the capability of outperforming a human, autonomous systems are still not flexible and intelligent enough to robustly handle all types of uncertainty (at least not to the degree of a human). Therefore, humans must continue to remain "in the loop" regardless of the immediate performance of their autonomous teammates.

In this paper, we propose a study with an experimental intervention that aims to demonstrate how the reliability of the autonomous agent affects the motivation, trust, and performance of their human teammates. Participants will play the team-based Predator-Prey game, where they team up with two autonomous agents trying to capture an autonomous prey. We then establish the following hypotheses:

- H_1: High autonomous teammate reliability leads to decreased human performance (i.e., complacency).
- H_2: High autonomous teammate reliability leads to reduced motivation.
- H_3: Human pre-disposition to complacency predicts observed performance, with a greater effect on low-performing autonomous teammates.

2. BACKGROUND

2.1 Human-Autonomy Teaming (HAT)

Due to the rising ubiquity of computation and autonomy as well as the growing interest in team-based interactions between human and machine, research in Human-Autonomy Teaming arose from the need to study behavior. Teamwork is defined as the "array of interconnected behaviors, cognitions, and attitudes that make coordinated and adaptive performance possible".[6] Traditionally, Salas et al. established the three-pronged approach to teamwork based on communication, coordination, and cooperation.[6] However, due to the automation's current growing abilities at following these requirements (e.g., communicating[10]), alternate approaches and their effects on human behavior must be investigated as automation progresses to have capabilities akin to another human.

Research has identified the three main roles autonomous agents can play: individual support, team support, or team member. Supporting roles have been largely researched in the context of passive and decision support systems, yet, agents as team members rely on a higher degree of specific interactions (e.g., predictability, expectations, team knowledge) which provide difficult research challenges.[11,12] In experimental approaches, autonomous agents have the capability to exchange information, communicate (albeit in a limited fashion), and verify and correct errors. Ideally, teams can be assisted in information retrieval, communication, monitoring, and planning.[13] However, in practice, due to the difficult rationalization and explainability of the autonomy, humans often end up frustrated with the expected performance and interactions of their autonomous teammates, leading to various performance issues such as tunnel vision, degraded situation awareness, and complacency.[12]

Applications of Human-Autonomy Teaming are already widely used in transportation (e.g., autonomous cars), military operations (e.g., high-value targeting, reconnaissance), and flight operations and dispatching,[12,14] to mention a few examples. Large advances are being made in developing software agents (e.g., decision support systems) and embodied agents (e.g., physical robots) in a variety of different contexts; agents that should be capable of understanding its and its team's tasks and conduct effective interaction with its teammates (human or otherwise).[15] However, due to either implementation limitations and external factors, the performance of autonomous agents can suffer, requiring a need not only to investigate how teams adapt to different levels of performance reliabilities, but also if they even detect failures at all.

2.2 Complacency, Automation, and Performance Degradation

Interaction with automated and decision support systems is a fundamental research landscape for human factors, with much research showing that support from automation can change human activity in unpredictable ways.[8] Decision support systems have long existed to assist in the decision making process, and as they become increasingly complex, they become difficult to predict and comprehend. While many designers respond by increasing transparency and customizability, this has still been shown to lead to a state of over-trusting, resulting in degradation of judgment[16] and knowledge.[17]

Complacency is defined as a "psychological state characterized by a low index of suspicion", possibly leading to operators not checking the system state enough for safe or optimal operation, assuming that "all is well".[8] Researchers mostly associate complacency with tasks that require supervisory control, leading to hampered, and even possibly fatal, human performance,[9] mainly due to a reduced frequency of checks in whether the machine is functioning correctly.[18] However, complacency is not exclusive to automation that requires supervision; autonomous agents have the capability of making a decision and executing without supervision. If the decision is trusted without regard by the human operator, it can lead to complacency through degraded performance.

Various studies have investigated different ways of counteracting complacency, as it is one of the predominant issues in automation.[19] Bagheri and Jamieson investigated the effect of different levels of reliability over time on failure detection[20] and how transparency serves to mitigate its resulting complacency and performance loss.[21] Salehi et al. used accountability in interactive control agents as a deterrent to complacency, with short term gains but slowing the decision-making process.[22] Bahner et al. suggest that training protocols can mitigate complacency.[23]

With AI-powered automation and autonomy becoming ubiquitous, we must begin considering complacent behavior in systems beyond monitoring-type interactions, as the role of autonomy evolves from being a passive bystander (e.g., recommendations, suggestions) to active executor and team member, akin to how we would treat and trust another human being.[3]

3. METHODOLOGY

In this section, we describe the methodology of the proposed study in detail. We aim to have participants play the Predator-Prey game in order to measure behavioral complacency throughout the task. We manipulate the reliability of the autonomous agents and measure the participant's performance throughout the task.

3.1 Predator-Prey Game

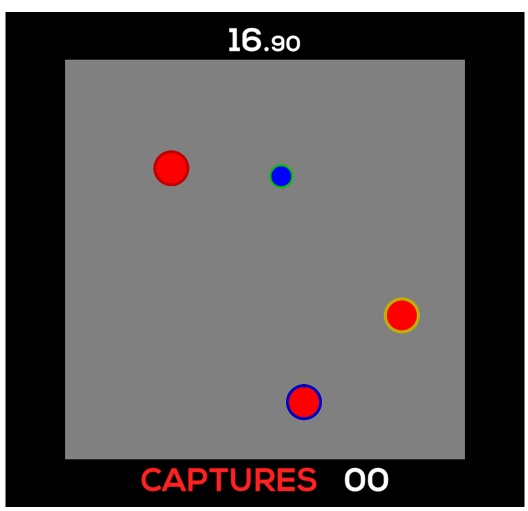

Figure 1. Screenshot of the Predator-Prey Game. 3 predators (red dots) team up to capture a prey (blue dot).

The experimental tool to provide the task and measure participants' performance is the Predator-Prey game (Fig. 1). Originally a testbed[24] to simulate and train machine learning models for cooperative asset (e.g., drones, tanks) maneuvers using OpenAI's gym toolkit,[25] we developed an environment in the Unity[26] game engine to allow human input and interaction. Participants are tasked with teaming up with two autonomous agents (predators) to "capture" (i.e., collide with) a third autonomous agent (prey), which is constantly evading in a continuous space. A continuous space serves to help our understanding on real-world pursuit tasks and strategy formation, as opposed to a simplified, discrete space.[24]

The task environment is comprised of a closed square arena of 2 m* of width and 2 m of height. The players in the Predator-Prey game move their circular avatar on a physics-based system by applying a force to their agent. The force is applied by human players through a joystick in an Xbox One controller, which translates to a vector. In order to give predators and prey an equal chance to succeed, as well as to encourage the emergence of coordination among predators, the predators were made slower than the prey. The predators had a maximum speed of 1 m/s and accelerated at a maximum rate of 3 m/s^2. The prey had a maximum speed of 1.3 m/s and accelerated at a maximum rate of 4 m/s^2. The mass of the players was set at 1 kg. The diameter of the players were 0.15 m and 0.1 m for predators and prey, respectively. Upon capture, the capturing predator and prey would knock each other back at an impulse force of 1 $\frac{m}{kg \cdot s}$ until losing all momentum (by reducing the absolute velocity at the rate of 0.25 m/s). Additionally, the prey is granted 0.5 seconds of invincibility after being captured, such that subsequent captures are not recorded if they are bumped by multiple predators at the same time.

The autonomous agents make decisions through a multi-agent deep deterministic policy gradient (MADDPG), a decentralized actor-centralized critic reinforcement learning algorithm which accounts for each agent's actions when searching for an optimal policy to execute (Fig. 2). A policy gradient serves appropriately due to the continuous action space of the Predator-Prey game.[27]

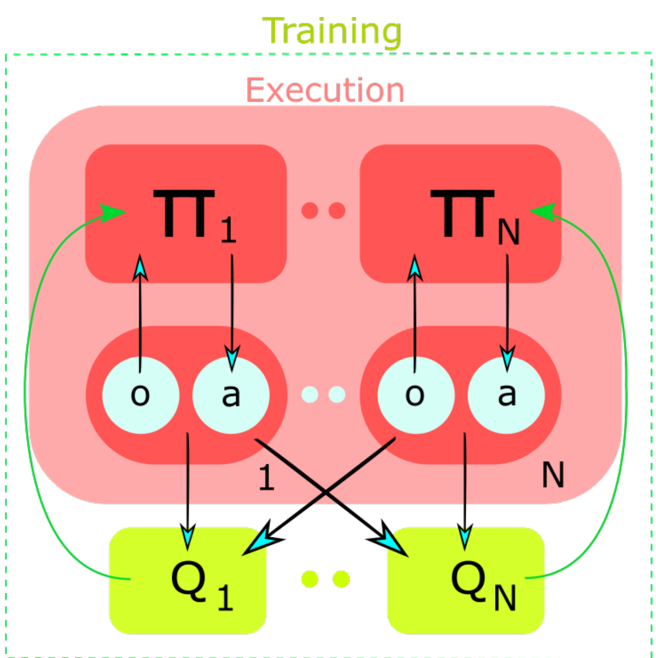

Figure 2. Overview of MADDPG's structure: a decentralized actor, centralized critic approach (adapted from Lowe et al.[27]). Policy π receives observations o (i.e., the state of the world) and outputs actions a (i.e., predator or prey movement), and then assigned reward Q (i.e., rewarded when capturing). It learns to associate the reward with the given action at that world state, building up the policy.

Figure 3 outlines the operational flow of the MADDPG algorithm. The algorithm itself is a neural network which takes 12 observations from the environment for each predator and 10 inputs for the prey as input, and

*Unity (or game engine) units are an arbitrary measure, but best practice is to equate it to an SI base unit.

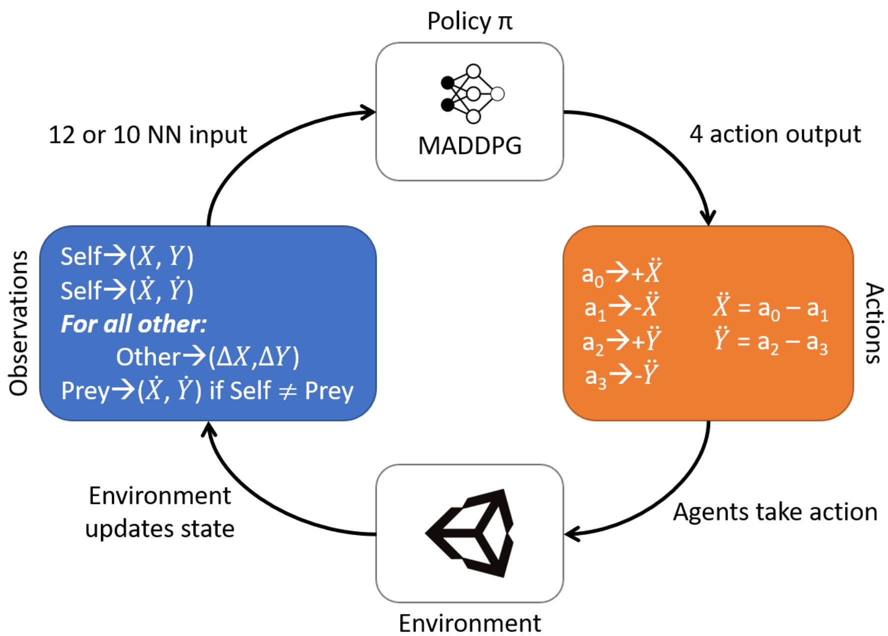

Figure 3. Overview of the MADDPG operational flow. MADDPG takes observations from the environment, and outputs the best actions according to its policy gradient. These actions are converted into behavior in the environment and updates its state for new observations.

outputs 4 actions for each agent. The observations fed into the neural net are the X and Y position of the current agent (2 inputs), the X and Y velocity of the current agent (2 inputs), the positional difference between the current agent and all other agents per axis (adding 6 inputs), and the X and Y velocity of the prey (2 inputs, ignored if the agent is the prey as velocity has already been accounted for). The algorithm then outputs 4 values ranging from 0 to 1 inclusive for each agent: the X-positive force (right), the X-negative force (left), the Y-positive force (up), and the Y-negative force (down). The actions are converted into a 2-dimensional force vector in the environment and applied to the corresponding agent, culminating in its movement.

We trained the autonomous agents with an MADDPG algorithm to learn how to play the Predator-Prey game through 200,000 episodes, with rewards assigned based on how many times a predator captured a prey (reward for predators, penalty for prey). Reinforcement learning algorithms search for the optimal policy during training, and in this scenario, an optimal policy is one such that a predator closes its distance to the prey, or strategizes with other predators to do so. Because humans are asynchronous to the agents' process and may act freely of this constraint (i.e., they may move opposite of an optimal policy, or not move at all), policies tend to fail whenever a member of the team behaves not according to the optimal policy (as the observation changes to a point in the gradient which may have been unexplored).[28] We attempted to counteract this issue by making single or combinations of agents in the predator team inactive during certain training episodes, so predators would learn to capture the prey without the intervention of the whole team, resulting in a model more robust to error.

In order to manipulate the reliability of the autonomous predators during the Predator-Prey game, we introduced noise according to condition to force errors in the model, as previous work has shown a relationship between the quality of observations and interdependence between agents' actions.[29] The agents' observations were perturbed with random values according to a logistic distribution. A logistic distribution gives us more values from the tails in comparison to a normal distribution, due to its higher kurtosis. The parameters were chosen arbitrarily and verified visually by interacting with the resulting agents. This allows a low, but reasonable percentage (logistic distribution with $\mu = 0$, $s = 0.03$, peturbation SD $= 0.054414$) of the observations to fall outside the model's expectations, leading the agents to occasionally behave erratically before recovering to the optimal policy.

To validate both the robust model (i.e., resistant to agent inactivity) and the noisy model (i.e., low reliability), we conducted a Kolmogorov–Smirnov (KS) test to justify 2 assertions: a) replacing the original model with our robust model would not cause changes in the observed complacent behavior by demonstrating they have congruent performance, and b) the different reliability models empirically perform differently, thus changes in the observed complacency behavior could possibly be attributed to the difference in reliability. Each trained model (original, robust, and noisy) was run for 50,000 episodes in testing (i.e., not learning any policies), and collected a distribution of their number of captures. A KS test found significant differences in both comparative distributions (original vs. robust: $D = 0.014$, $p < 0.001$; robust vs. noisy: $D = 0.0278$, $p < 0.001$). While this test does not support assertion *a*, we do find that the effect size for this contrast is less than half of that for assertion *b*. Further, in 70% of the episodes, the predators never captured the prey, providing a large amount of data and increasing the power of the KS test to find significant differences. In light of this, we conducted an additional KS test where we excluded episodes with 0 captures in the analysis. Figure 4 outlines the empirical cumulative distribution functions for the number of captures of the three models. For assertion *a*, the KS test found no significant difference in captures between the original and robust models ($D = 0.008$, $p = 0.68$), thus it is replaceable to account for unwanted human behavior. For assertion *b*, the KS test found significant difference in captures between the robust and noisy models ($D = 0.078$, $p < 0.001$), thus observed behavior during the intervention can be attributed to the difference in reliability. These results were replicated with a Zero-Inflated Poisson Regression, which validates both assertion *a* (original vs. robust: $b = 0.005$, $p = 0.507$, difference in log-likelihood with null model: 10.7, $p < 0.001$) and *b* (robust vs. noisy: $b = -0.125$, $p < 0.001$, difference in log-likelihood with null model: 127.06, $p < 0.001$). As per our expectations, the original and robust model did not have a significant difference, whereas the noisy model differed from the robust model.

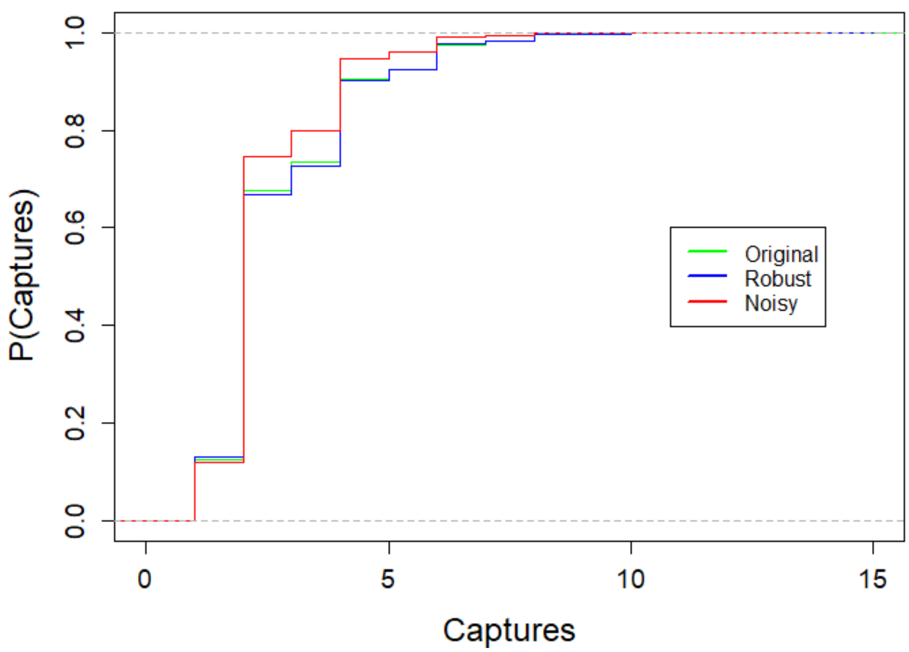

Figure 4. Empirical cumulative distribution functions (ECDFs) for the original, robust, and noisy MADDPG models. Episodes with 0 captures were dropped from the model due to zero-inflation.

3.2 Experimental Design

The Predator-Prey game was designed to simulate the interactions between operators engaged in a pursuit-evasion task. This task sets predators and prey with competing, mutually exclusive goals, where the result of the task highly depends on the collaborative and individual performances of each entity. Models of predator-prey pursuit behavior are useful for identifying coordination strategies, and they serve as an intuitive and archetypal collaboration task for participants.[30]

Participants are tasked to play 20 rounds of the Predator-Prey game for 30 seconds each. As mentioned in section 3.1, the participant takes on the role of a predator, assisted by 2 autonomous teammates to capture an autonomous prey. Participants are allowed 2 practice rounds before the actual recorded runs begin. All predators begin at least 0.5 m away from the prey, as to not provide any advantageous position to either team (i.e., predators will not begin by cornering the prey).

Before and after the intervention, participants completed a battery of questions to measure various subjective characteristics pertaining to technology and automation. These questions consisted of validated scales to measure motivation, complacency potential, trust propensity, perception of trust, and perceived workload. The inclusion of these scales serves an exploratory purpose to find a relationship between personal characteristics and performance with respect to the experimental intervention (i.e., autonomous agent's performance). Surveys were modified and adapted to measure the criteria of interest. Figure 5 outlines the proposed experimental flow in 5 stages. Table 1 outlines the order of the surveys before and after intervention.

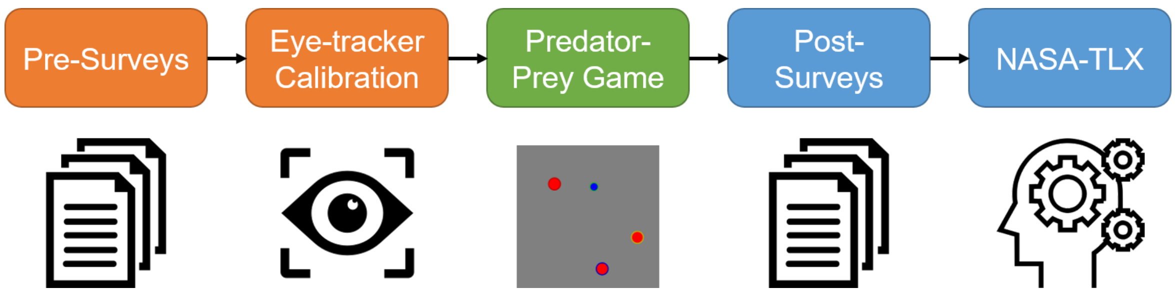

Figure 5. Proposed experimental flow.

An a priori power analysis was conducted and indicates that given a significance level α of 0.05, the required sample size n for a demonstrable effect is either 128 (false negative probability $\beta = 0.2$, effect size $f = 0.25$) or 84 (false negative probability $\beta = 0.05$, effect size $f = 0.4$).

The independent variable of interest is the reliability of the agent: high or low. The reliability was persistent throughout the entire experiment and was manipulated between-subjects.

3.3 Measures

3.3.1 Behavioral Observations

An eye tracker will be used to capture the participants' focus and fixation points throughout the trials. Gaze metrics have been widely used to study cognition in multiple studies, revealing insights in how humans acquire domain knowledge and complete tasks.[31] Throughout the task, multiple potential focal points exist, such as the participant's avatar, the prey, and the central point of a polygon with each player in the game as its vertices. The frequency of gazing these areas of interest will serve as the quantification of attention and provide insight into complacent behavior.

Individual participant performance and team performance will be recorded, quantified by the amount of captures the predators achieve. Performance is the main dependent variable analogous to mission success in real-life scenarios. All positional, input, and eye tracking data will be retained for replay and re-simulation. This results in multiple time series which can be subject to time series analysis for insight on team-based behavior. The data points are sampled at 60 Hz.

3.3.2 Surveys

The Intrinsic Motivation Inventory (IMI) presents a 22 item version which focuses on task evaluation, measuring interest and enjoyment, perceived choice, perceived competence, and pressure and tension.[32] Designed to be reworded to contextualize the inventory for the task, questions were modified to make them relevant to the Predator-Prey game (e.g., "I felt pretty skilled when cooperating with others" instead of "I felt pretty skilled at this task"). The IMI was given before and after the intervention to measure changes in motivation, which has been known to accurately predict performance and engagement in game-based tasks.[33]

The Automation-induced Complacency Potential (AICP) scale measures a participant's tendency toward sub-optimal monitoring patterns through 2 factors: workload alleviation and frequency of monitoring.[34] The AICP was given before the intervention, and it should relate to their propensity to trust autonomy and their perceived trust of the autonomous teammates. Although the AICP measures complacency through automation, there is no scale for interaction with autonomous agents, and thus it serves as an initial point of data collection and inference.

The adapted Propensity to Trust Technology (aPTT) scale measures a participant's general tendency to trust technology, and it was modified in this study by including language that specifically referred to "automated agents" rather than technology.[35] According to Jessup et al., using the specific language of "automated agents" allows the measure to predict behavioral trust. This is a measure of trust before the participant interacts with the autonomous agents.

The Trust in Automated Systems (TAS) scale measures how trustworthy the participants perceived the system that they just interacted with – in our case, the autonomous teammates[36] – to be. The questions in this scale were rephrased with the autonomous teammates as the object of reference. We predict the observed trust measured by this scale to be related to the previous aPTT.

The NASA Task Load Index (NASA-TLX) is a widely-used and strongly validated tool to measure perceived workload throughout a task.[37] Since complacency might relate to the amount of workload a participant perceives, we expect to find a correlation between agent reliability and perceived workload. Additionally, any variance demonstrated from predicting complacent behavior using reliability could be clarified using workload.

All survey items are detailed in Appendix A.

Table 1. Surveys and measurements utilized before and after interventions.

Before Intervention	**After Intervention**
IMI (Ryan and Deci, 1982)	IMI (Ryan and Deci, 1982)
AICP (Merritt et al., 2019)	TAS (Jian et al., 2000)
aPTT (Jessup et al., 2019)	NASA-TLX (Hart and Staveland, 1988)

4. DISCUSSION

The goal of this work is to facilitate the development and design of autonomous agents by providing insight on the effect of agent reliability on the performance of their human teammates. We approach this with three hypotheses, reiterated below.

- H_1: High autonomous teammate performance leads to decreased human performance (i.e., complacency).
- H_2: High autonomous teammate performance leads to reduced motivation.
- H_3: Human pre-disposition to complacency predicts observed performance, with a greater effect on low-performing autonomous teammates.

With rapid advances in the capabilities of autonomous agents and their ability to mimic human behavior and surpass performance, there is a need to investigate the limits of human cooperation with autonomy, with the aim of allowing intuitive teamwork between them.

We use task performance (i.e., team and individual captures) as our main affected metric by agent reliability. We hypothesize that as the agent is more reliable and skillful, human operators begin offloading their cognitive efforts to the automation, resulting in reduced performance and/or increased complacency throughout the task. H_1 is tested by using task performance as a metric; we can conduct a t-test or Mann-Whitney test based on normality to establish significant differences between participant performance interacting with reliable or unreliable agents.

The battery of questions before and after intervention serve to answer H_2 and H_3. The IMI is given before and after the intervention, and our sub-scales of interest are interest/enjoyment (as a metric of intrinsic motivation), and pressure/tension (analogous to complacency and perceived workload). With aggregated scores and a Wilcoxon test (due to its ordinality and paired data points before and after intervention), it should indicate any changes in motivation after interacting with the autonomous agents. We expect to see greater drops in motivation in the low-reliability condition than in the high-reliability condition, per H_2.

The AICP and aPTT are predictors of complacent behavior and trust, measured in the intervention by the Predator-Prey game and the TAS, respectively. Intuitively, a person who is more likely to be complacent and has a high level of trust in automation will show a higher degree of behavioral complacency (i.e., lower performance in the Predator-Prey game). Thus, we expect (per H_3) that predisposition to complacency predicts performance in a team with autonomous teammates, with a higher degree of accuracy if the teammates are unreliable. This hypothesis will be tested with a multiple linear regression or a generalized linear model with AICP and aPTT as predictors, and TAS as a co-variate. Additionally, a mixed model using reliability as a fixed effect and participants' subjective tendencies as a random effect may help us to model the observed performance's variance more clearly.

Eye-tracking data has long been used in human factors to record focus points and fixations,[31] and it can reveal attention allocation and strategies throughout the Predator-Prey game. Ideally, a participant's gaze should follow a trajectory between their own controlled predator and the prey, with occasional glances to the general vicinity of the autonomous teammates and the status of the task (i.e., the capture count and the timer). A complacent participant would instead disengage from the task, and their gaze would have degraded and delayed tracking to the object of interest (i.e., the prey). The reduced frequency of monitoring the behavior of the autonomous teammates would serve as the quantification of complacent behavior.

For an agent to identify and participate in a non-explicit strategy initiated by a human teammate, they must be able to detect different strategies from observable data. Ultimately, the aim is to allow intuitive interactions between humans and autonomy, and in the case of continuous task spaces, agents should be able to adapt to strategies bottom-up (i.e., recognizing units of behavior as part of a larger strategy), while preventing performance degradation in their human teammates. This is important because, while strategies can be implemented top-down (i.e., declared conceptually, then implementing the small behaviors), they are often fluid, implicit, and emergent. Recent work has demonstrated that univariate time series methods can be used to detect ground-truth changes in strategy (as defined by agent policy) from summaries of an all-agent team's spatial positions.[30] Positional data from the Predator-Prey game can allow a first attempt at identifying group strategies between humans and autonomous agents in a team-based task.

In the context of artificial intelligence and machine learning development, a central issue which remains is the understandability of the final behavior of a trained autonomous agent. With respect to our implementation, neural networks are inherently "black boxes", and thus the internal representations of the environments and policies it holds cannot be known. As this affects understanding of the behavior, it can be difficult to detect when autonomy is engaging in cooperative behavior. To this avenue, recent work has explored using convergent cross mapping[38] and ergodic spatial distributions[39] in order to gain insight of the behavior between heterogeneous teams and establish whether they are actively cooperating versus working independently but with a shared goal.

A novel approach to automation complacency in multi-agent systems would be to view it from the lens of social loafing.[40] Social loafing is formally defined as the "reduction in motivation and effort when individuals work collectively compared with when they work individually or coactively".[41] Performance degradation due to complacency serves akin to the penalties incurred due to social loafing, as the human does not need to invest as much effort in the task due to the perceived ability of the autonomous agents. Disengagement from the task is ultimately an example of team cognition degrading, due to the influence of the autonomous agents to their human counterparts. A future avenue of work possibly entails adapting solutions to social loafing to the domain of HAT and quantifying their effectiveness.

4.1 Limitations

This proposal provides a step forward in investigating complacency and performance to design effective autonomous teammates, however, many challenges remain. The Predator-Prey game only represents a single task in a wide variety of team-based continuous tasks that may represent the real world. Modifying goal conditions, simulation parameters such as number of teammates or agent velocities, or the environment itself can yield a different scenario where generalization of performance becomes challenging. However, it serves as a starting point for the effects of complacency on performance in multi-domain operations, particularly when the environment can be uncertain.

Complacency still remains loosely defined in practice, beyond attempts to provide an empirical model for its measurement based on attention sampling.[18] Although complacency is defined by task, the commonality between the definitions lie between monitoring the automation and demonstrated task performance. We predict in this study that demonstrated performance and the frequency in which the autonomous teammates are monitored are a proxy for complacent behavior.

Investigating automation reliability in a multi-agent continuous task is a novel approach to the human-autonomy teaming research, and although we have framed demonstrated task performance and related co-variates in terms of complacency, the possibility remains that other extraneous factors describe our results instead. The variance of our resulting generalized linear model should indicate performance not explained by agent reliability or any of the subjective predispositions to complacent behavior.

APPENDIX A. SURVEYS AND QUESTIONNAIRES

The following are the used surveys before and after the intervention. Items denoted with R are reverse scored.

A.1 Intrinsic Motivation Inventory (Ryan and Deci 1982)

Interest/enjoyment: 1, 5, 8, 10, 14 (R), 17, 20
Perceived competence: 4, 7, 12, 16, 22
Perceived choice: 3, 11 (R), 15, 19 (R), 21 (R)
Pressure/tension: 2 (R), 6, 9 (R), 13, 18

1. While I was cooperating with AI I was thinking about how much I enjoyed it.
2. I did not feel nervous at all about doing cooperating with AI.
3. I felt that it was my choice to cooperate with AI.
4. I think I am pretty good at tasks that involve cooperating with AI.
5. I find cooperating with AI very interesting.
6. I feel tense while cooperating with AI.
7. I think I did pretty well at cooperating with AI, compared to others.
8. Cooperating with AI is fun.
9. I feel relaxed while cooperating with AI.
10. I enjoy cooperating with AI very much.
11. I don't really have a choice when cooperating with AI.
12. I am satisfied with cooperating with AI.
13. I am anxious when cooperating with AI.
14. I think cooperating with AI is very boring.
15. I feel like doing what I want to do while I cooperate with AI.
16. I feel pretty skilled when cooperating with AI.
17. I think cooperating with AI is very interesting.

18. I feel pressured while cooperating with AI.
19. I feel like I have to cooperate with AI.
20. I would describe cooperating with AI as very enjoyable.
21. I cooperate with AI because I have no choice.
22. After cooperating with AI for awhile, I feel pretty competent.

A.2 Automation-induced Complacency Potential Scale (Merrill et al. 2019)

Alleviating workload: 1, 2, 3, 4, 6
Monitoring: 5 (R), 7, 8 (R), 9, 10

1. When I have a lot to do, it makes sense to delegate a task to automation.
2. If life were busy, I would let an automated system handle some tasks for me.
3. Automation should be used to ease people's workload.
4. If automation is available to help me with something, it makes sense for me to pay more attention to my other tasks.
5. Even if an automated aid can help me with a task, I should pay attention to its performance.
6. Distractions and interruptions are less of a problem for me when I have an automated system to cover some of the work.
7. Constantly monitoring an automated system's performance is a waste of time.
8. Even when I have a lot to do, I am likely to watch automation carefully for errors.
9. It's not usually necessary to pay much attention to automation when it is running.
10. Carefully watching automation takes time away from more important or interesting things.

A.3 Adapted Propensity to Trust in Technology (Jessup et al. 2019, original: Schneider et al. 2017)

1. Generally, I trust automated agents.
2. Automated agents help me solve many problems.
3. I think it's a good idea to rely on automated agents for help.
4. I don't trust the information I get from automated agents. (R)
5. Automated agents are reliable.
6. I rely on automated agents.

A.4 Trust in Automated Systems (Jian et al. 2000)

1. The agent is deceptive. (R)
2. The agent behaves in an underhanded manner. (R)
3. I am suspicious of the agent's intent, action, or outputs. (R)
4. I am wary of the agent. (R)
5. The agent's actions will have a harmful or injurious outcome. (R)
6. I am confident in the agent.
7. The agent provides security.
8. The agent has integrity.
9. The agent is dependable.
10. The agent is reliable.
11. I can trust the agent.
12. I am familiar with the agent.

Acknowledgments

This research was sponsored by the Army Research Laboratory. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

[1] Arrow, H., McGrath, J. E., and Berdahl, J. L., "Small groups as complex systems: Formation, coordination, development, and adaptation," (2000).

[2] Fiore, S. M. and Wiltshire, T. J., "Technology as teammate: Examining the role of external cognition in support of team cognitive processes," *Frontiers in Psychology* **7**, 1531 (2016).

[3] Kessler, T., Stowers, K., Brill, J., and Hancock, P., "Comparisons of human-human trust with other forms of human-technology trust," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **61**(1), 1303–1307 (2017).

[4] Barton, S. L. and Asher, D., "Reinforcement learning framework for collaborative agents interacting with soldiers in dynamic military contexts," in [*Next-Generation Analyst VI*], Hanratty, T. P. and Llinas, J., eds., **10653**, 14 – 27, International Society for Optics and Photonics, SPIE (2018).

[5] Demir, M., McNeese, N. J., and Cooke, N. J., "Team communication behaviors of the human-automation teaming," in [*2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*], 28–34 (March 2016).

[6] Salas, E., Wilson, K. A., Murphy, C. E., King, H., and Salisbury, M., "Communicating, coordinating, and cooperating when lives depend on it: Tips for teamwork," *The Joint Commission Journal on Quality and Patient Safety* **34**(6), 333 – 341 (2008).

[7] Langner, R. and Eickhoff, S., "Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention," *Psychological bulletin* **139**, 870–900 (07 2013).

[8] Parasuraman, R. and Manzey, D. H., "Complacency and bias in human use of automation: An attentional integration," *Human Factors* **52**(3), 381–410 (2010). PMID: 21077562.

[9] Parasuraman, R. and Riley, V., "Humans and automation: Use, misuse, disuse, abuse," *Human factors* **39**(2), 230–253 (1997).

[10] Demir, M., McNeese, N. J., Cooke, N. J., Ball, J. T., Myers, C., and Frieman, M., "Synthetic teammate communication and coordination with humans," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **59**(1), 951–955 (2015).

[11] Sukthankar, G., Shumaker, R., and Lewis, M., "Intelligent agents as teammates," in [*Theories of Team Cognition: Cross-Disciplinary Perspectives*], Fiore, S. and Letsky, M., eds., 313 – 343, Routledge Academic (2011). Please access chapter through related URL.

[12] Chen, J. and Barnes, M., "Human–agent teaming for multirobot control: A review of human factors issues," *Human-Machine Systems, IEEE Transactions on* **44**, 13–29 (02 2014).

[13] Sycara, K., "Integrating agents into human teams," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **46**(3), 413–417 (2002).

[14] Ho, N., Johnson, W., Panesar, K., Wakeland, K., Sadler, G., Wilson, N., Nguyen, B., Lachter, J., and Brandt, S., "Application of human-autonomy teaming to an advanced ground station for reduced crew operations," in [*2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*], 1–4 (Sep. 2017).

[15] Demir, M., McNeese, N. J., and Cooke, N. J., "The evolution of human-autonomy teams in remotely piloted aircraft systems operations," *Frontiers in Communication* **4**, 50 (2019).

[16] Wright, J. L., Chen, J. Y. C., Barnes, M. J., and Hancock, P. A., "The effect of agent reasoning transparency on automation bias: An analysis of response performance," in [*Virtual, Augmented and Mixed Reality*], Lackey, S. and Shumaker, R., eds., 465–477, Springer International Publishing, Cham (2016).

[17] Rodriguez, S. S., O'Donovan, J., Schaffer, J. A., and Höllerer, T., "Knowledge complacency and decision support systems," in [*2019 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*], 43–51 (April 2019).

[18] Bagheri, N. and Jamieson, G., "A sampling model to ascertain automation-induced complacency in multitask environments," *International Federation for Information Processing Digital Library; Human Error, Safety and Systems Development;* **152** (01 2004).

[19] Parasuraman, R., Molloy, R., and Singh, I. L., "Performance consequences of automation-induced'complacency'," *The International Journal of Aviation Psychology* **3**(1), 1–23 (1993).

[20] Bagheri, N., Jamieson, G. A., et al., "Considering subjective trust and monitoring behavior in assessing automation-induced "complacency."," *Human performance, situation awareness, and automation: Current research and trends* **1**, 54–59 (2004).

[21] Bagheri, N. and Jamieson, G. A., "The impact of context-related reliability on automation failure detection and scanning behaviour," in [*2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*], **1**, 212–217, IEEE (2004).

[22] Salehi, P., Chiou, E. K., and Wilkins, A., "Human-agent interactions: Does accountability matter in interactive control automation?," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **62**(1), 1643–1647 (2018).

[23] Bahner, J. E., Hüper, A.-D., and Manzey, D., "Misuse of automated decision aids: Complacency, automation bias and the impact of training experience," *International Journal of Human-Computer Studies* **66**(9), 688 – 699 (2008).

[24] Asher, D. E., Zaroukian, E., and Barton, S. L., "Adapting the predator-prey game theoretic environment to army tactical edge scenarios with computational multiagent systems," (2018).

[25] Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., and Zaremba, W., "Openai gym," (2016).

[26] "Unity real-time development platform," (2020).

[27] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., and Mordatch, I., "Multi-agent actor-critic for mixed cooperative-competitive environments," (2017).

[28] Moosavi-Dezfooli, S., Fawzi, A., and Frossard, P., "Deepfool: a simple and accurate method to fool deep neural networks," *CoRR* **abs/1511.04599** (2015).

[29] Derrik E. Asher, Michael Garber-Barron, S. S. R. E. Z. and Waytowich, N. R., "Multi-Agent Coordination Profiles Through State Space Peturbations," (In Press).

[30] Zaroukian, E., Rodriguez, S. S., Barton, S. L., Schaffer, J. A., Perelman, B., Waytowich, N. R., Hoffman, B., and Asher, D. E., "Algorithmically identifying strategies in multi-agent game-theoretic environments," in [*Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*], Pham, T., ed., **11006**, 400 – 412, International Society for Optics and Photonics, SPIE (2019).

[31] Eckstein, M. K., Guerra-Carrillo, B., Singley, A. T. M., and Bunge, S. A., "Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development?," *Developmental Cognitive Neuroscience* **25**, 69 – 91 (2017). Sensitive periods across development.

[32] Ryan, R. M. and Deci, E. L., "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.," *American Psychologist* **55**(1), 68–78 (2000).

[33] Ryan, R. M., Rigby, C. S., and Przybylski, A., "The motivational pull of video games: A self-determination theory approach," *Motivation and Emotion* **30**, 344–360 (nov 2006).

[34] Merritt, S. M., Ako-Brew, A., Bryant, W. J., Staley, A., McKenna, M., Leone, A., and Shirase, L., "Automation-induced complacency potential: Development and validation of a new scale," *Frontiers in Psychology* **10**, 225 (2019).

[35] Jessup, S. A., Schneider, T. R., Alarcon, G. M., Ryan, T. J., and Capiola, A., "The measurement of the propensity to trust automation," in [*Virtual, Augmented and Mixed Reality. Applications and Case Studies*], Chen, J. Y. and Fragomeni, G., eds., 476–489, Springer International Publishing, Cham (2019).

[36] Jian, J.-Y., Bisantz, A. M., and Drury, C. G., "Foundations for an empirically determined scale of trust in automated systems," *International Journal of Cognitive Ergonomics* **4**(1), 53–71 (2000).

[37] Hart, S. G. and Staveland, L. E., "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in [*Human Mental Workload*], Hancock, P. A. and Meshkati, N., eds., *Advances in Psychology* **52**, 139 – 183, North-Holland (1988).

[38] Barton, S. L., Waytowich, N. R., Zaroukian, E. G., and Asher, D. E., "Measuring collaborative emergent behavior in multi-agent reinforcement learning," *CoRR* **abs/1807.08663** (2018).

[39] Derrik E. Asher, Erin Zaroukian, B. P. J. P. R. F. B. H. S. S. R., "Multi-Agent Collaboration with Ergodic Spatial Distributions," (In Press).

[40] Latané, B., Williams, K. D., and Harkins, S. G., "Many hands make light the work: The causes and consequences of social loafing.," (1979).

[41] Karau, S. and Williams, K., "Social loafing: A meta-analytic review and theoretical integration," *Journal of Personality and Social Psychology* **65**(4), 681–706 (1993).