CHAOPT: A Testbed for Evaluating Human-Autonomy Team Collaboration Using the Video Game Overcooked!2

Justin Bishop[1], Jaylen Burgess[1], Cooper Ramos[1], Jade B. Driggs[1], Chad C. Tossell[1], Tom Williams[2],
Elizabeth Phillips[1], Tyler H. Shaw[3] & Ewart J. de Visser[1]
United States Air Force Academy[1], Colorado School of Mines[2], George Mason University[3]

***Abstract* – This paper introduces a new testbed called Cooking with Humans and Autonomy in Overcooked!2 for studying Performance and Teaming (CHAOPT). A validation study was conducted to examine the viability of Overcooked!2 as a research platform to explore teamwork and communication in human-autonomy teams. Unique measures derived from this platform such as productive chef actions (PCA), team expertise score and chef role contribution (CRC) distinguished performance between levels and players. Our findings demonstrate that we can derive meaningful team process, performance and communication measures and that the interactions within Overcooked!2 meet the requirements of psychological fidelity of teaming research.**

***Index Terms* – Autonomous agents, human-machine teaming, artificial intelligence, communication, trust**

INTRODUCTION

Low cost simulation methods including commercial-off-the-shelf games (COTS) have been popular tools for researchers studying human-robot interaction for several years as they offer many benefits to the research community [1]. COTS games are often significantly more cost effective, easy to acquire, and accessible to users than traditional simulation building software. Game-play in COTS is also purposefully designed to be fun, engaging, and immersive for users; thus, laboratory participants often enjoy participation which has the potential to yield more and better data than other simulation methods [1]. For research projects that require short execution times, like capstone projects led by senior undergraduate students [2], the engaging qualities of COTS make the project attractive to both research participants and research assistants.

Recent research has uncovered that when working in teams, communication between humans and autonomous agents is not as effective, efficient, or useful as between humans and humans working in teams [3]. As such, new methods may be needed to study and thereby improve such communication processes when humans and autonomous work together. We propose that the game Overcooked!2 may be an ideal COTS to explore human-autonomy team processes like communication. The game may have an adequate level of *psychological fidelity*--or how well a simulation replicates the same necessary and sufficient psychological cues and cognitive processes that occur while engaged in the real-world tasks the system is attempting to model [1,4]. The purpose of this paper is to report on the psychological fidelity afforded by and utility of the game Overcooked2! in the study of human-agent and human-human team process and performance.

THE CHAOPT TESTBED

This paper introduces a new testbed known as CHAOPT: **C**ooking with **H**umans and **A**utonomy in **O**vercooked! 2 for studying **P**erformance **T**eaming. This testbed is a product of a unique collaboration between Air Force Research Laboratory's (AFRL) Gaming Research Integration for Learning Laboratory (GRILL) East division, located at Wright-Patterson Air Force Base, and the GRILL West division residing at the U.S. Air Force Academy. The purpose of creating this tool is to develop the collaborative teamwork capabilities of autonomous agents by studying human-human and human-autonomy teamwork under chaotic conditions created by high uncertainty, volatility and stress. Overcooked!2 is an ideal choice for the study of this type of teamwork for several reasons. First, the chaotic nature of the game continually forces new strategies and thus necessitates frequent communication between agents. Second, the focus on collaborative teamwork and communication provides a challenging litmus test for an artificially intelligent agents' capabilities, which have traditionally focused on competitively beating human players in games such as Chess, Go and Starcraft [5,6], as opposed to working collaborative with human players. Lastly, the hellish kitchens designed in Overcooked!2 approximate a level of chaos that can characterize operational environments such as battlefields in military settings or emergency rooms in medical settings, among others. Thus, CHAOPT can be used to quantify and validate team performance metrics that are of interest to those studying teamwork between humans and autonomous agents or between humans and other humans.

I. Game Elements in Overcooked!2 that Promote Teamwork

The goal of Overcooked2! is for a team of up to 4 players playing as chefs in a kitchen to complete as many food orders as possible in a short amount of time (see Figure 1). In this effort, players are required to navigate the kitchen environment, prepare and cook the individual ingredients to complete orders, wash dishes and deal with distracting obstacles in the environment, all while attempting to deliver

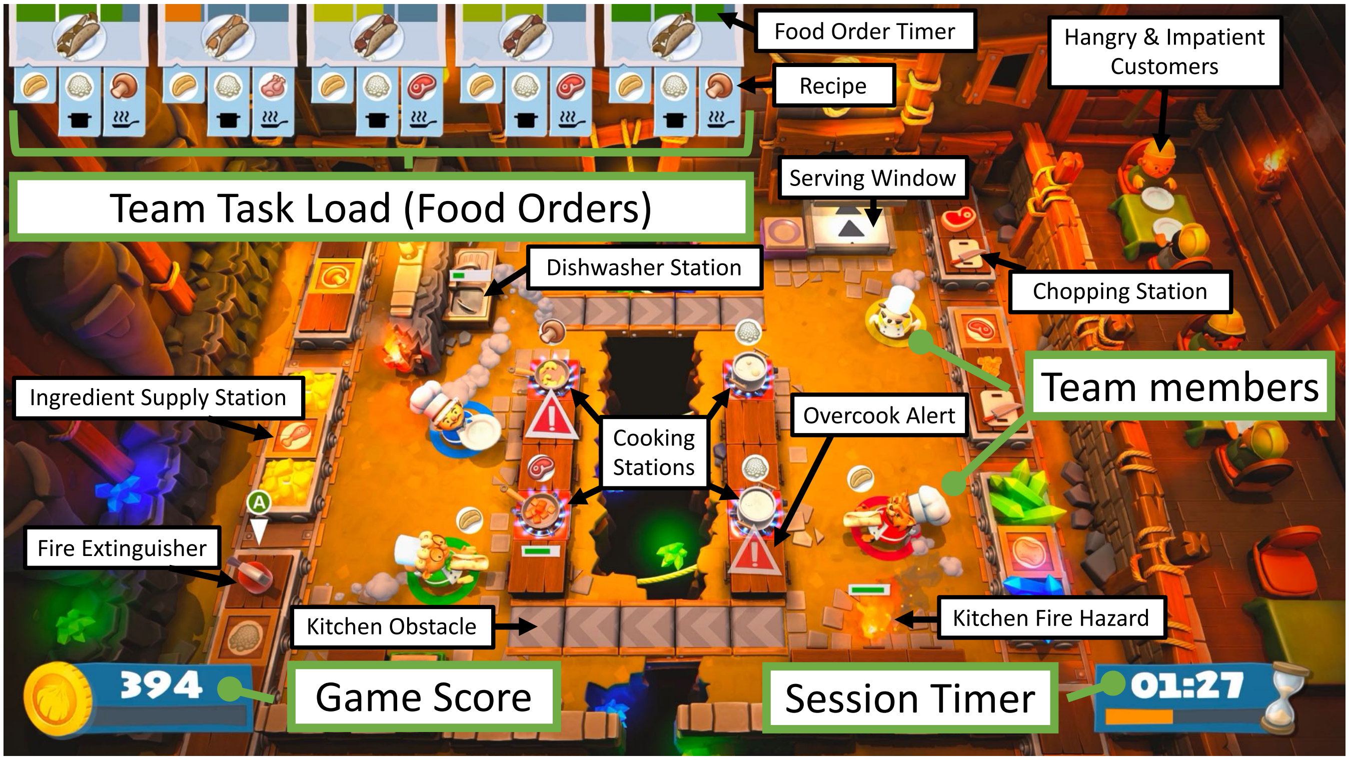

FIGURE 1. A HELLISH KITCHEN IN OVERCOOKED!2. LABELS INDICATE THE CRITICAL GAME ELEMENTS THAT RELATE TO TEAMWORK

the food orders on time in the correct sequence. Once orders are delivered, players are awarded points as well as tips based on speed and priority of delivery. Players lose points if orders are not fulfilled in time or served out of order. Much of gameplay strategy revolves around efficiently managing time. The session timer indicates the overall time that is available to complete as many orders as possible, while individual food orders have their own timers which directly determine the size of the tips players can accrue. The faster a team delivers a given food order the more points the team can accumulate. Lastly, dishes or individual ingredients require a certain amount of time to cook. As cooking gets closer to completion, a beeping alarm begins to sound. If a given dish or ingredient is not taken off the burner in time, the dish risks being overcooked which can result in a fire that can spread across and burn down the remainder of the kitchen and must then be extinguished with a fire extinguisher.

Overcooked further promotes communication and teamwork by forcing role-asymmetry through clever kitchen design. It is possible to assign roles for tasks neatly in the beginning of each level, but commonly as each session progresses, a number of hazards might occur which can disrupt the workflow of the team, thus requiring players to renegotiate roles and responsibilities during the course of gameplay. For example, players can easily fall into the precipitous chasms that are present in many kitchens, resulting in a 5-second delay before that character can respawn. Random fires may ignite that can form obstacles for players. Or, the spatial layout of many kitchens will suddenly and dynamically change. These changes in the environment put additional pressure on the team to constantly switch roles to maintain performance.

II. Testbed Player Setup with Wizard of Oz (WoZ)

A Wizard of Oz (WoZ) paradigm was used to enable the research team to facilitate gameplay between co-located and/or remote agents: humans and humans simulating autonomous players. WoZ is a control paradigm in which an experimenter remotely operates a robot or other artificial agent in order to simulate the autonomous capabilities or intelligences needed for a given context, study, or evaluation [7]. Two Nintendo Switch video game consoles were used to facilitate gameplay between participants and an experimental confederate or the confederate simulating an autonomous player. The Nintendo Switch is a small console that can be played either as a handheld, portable device or as a home console, by connecting it to a digital monitor via an HDMI cable. In the co-located setup, the participant and experimenter played with two separate controllers in front of a TV. In the remote conditions, the participant played in front of the tv while the experimenter played with the Switch in the portable format in another room over a wireless internet connection (see Figure 2).

To facilitate video recording of participant and confederate gameplay, we used a screen capture device called a Game Capture Card, that records console gameplay and stores the digital recordings on a storage device. Two mobile phones set to speaker phone and connected to wireless internet allowed participants to communicate with confederates while team members were dislocated. To record the audio communications between participants and confederates over the mobile phone connection, we connected a tabletop microphone to the Game Capture Card. This allowed us to simultaneously record gameplay and

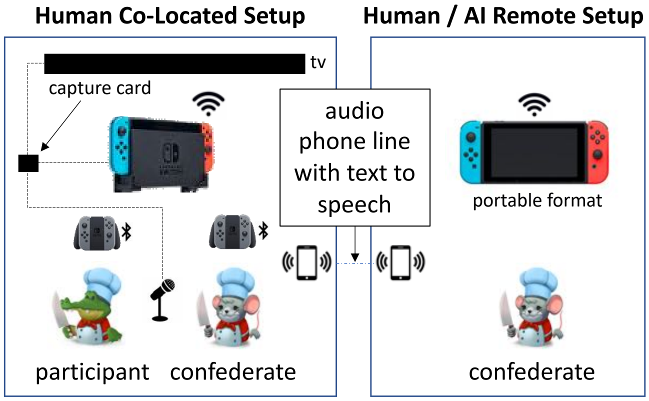

FIGURE 2. SCHEMATIC OF THE EXPERIMENTAL WIZARD OF OZ SETUP

participant and confederate communications and synchronize the audio and video recordings. Simulated autonomous player communications were performed by the confederate speaking through a voice modulation app.

III. Measures

To assess the utility of the CHOAPT testbed, we employed several existing subjective measures of affective team performance in a human-subjects study. Additionally, we devised several behavioral metrics afforded by the game which can be used to evaluate team process and performance (see Table 1).

VALIDATION STUDY

To assess the viability of using Overcooked!2 as a testbed for human-autonomy teaming research, we designed a human-subjects study to collect data for measure validation and to provide a preliminary test of potential experimental conditions that could be executed using the CHOAPT testbed. This study had participants attempt to complete six levels of Overcooked! 2 on the Nintendo Switch and were either instructed that they were playing with a co-located human confederate, a remotely located human confederate, or a remotely located autonomous partner, which was also played by a human confederate in a WoZ paradigm. The confederate player was absent from the gaming lab for all but the co-located condition to avoid the detection of the WoZ paradigm.

METHODS

I. Participants

Disruptions to in-person data collection due to the spread of COVID-19 resulted in only reporting on our pilot participant data collection in this paper. In total, nine participants (4 females, with ages ranging from 18-22, $M = 20.1$, $SD = 2.03$) participated in the Overcooked validation testing. Nine Overcooked teams were formed by allocating each novice participant to play with one of three available skilled confederate players, who played in supportive/helpful style. All participants were recruited from the United States Air Force Academy and received extra credit in their introductory level behavioral science course for completing the study.

II. Task Paradigm

Participants were instructed to play through six levels and to earn three stars for each level before passing to the next one. Each additional star earned corresponded to a higher game score. This design was purposefully chosen because when re-attempting levels novel strategies are often required, which stimulates communication between team members. Evaluating communication was our primary objective in assessing the testbed. Each level varied in its level of difficulty depending on recipe complexity, number of different recipes, and kitchen layout.

TABLE 1. SUBJECTIVE AND BEHAVIORAL MEASURE DEFINITIONS

Measure	Description
Team Perception	Scales (1-7) that assessed team trust, affect, cohesion, collaboration, effectiveness and role clarity [8]
Game Experience	Scales (1-7) that assessed usability, enjoyment and social experience [9]
Game Score	Game reported score calculated as follows: (served orders * order value) + tips – failed orders
Team Expertise Score (TSE)	Game score divided by the listed 4-star score for two players (cite). Achieving the 4-star score indicates mastery for a level. TSE indicates relative expertise.
Productive Chef Actions (PCA)	The sum of all cumulative actions that it takes to put together a food order: one action for each ingredient that was prepared (picking up, chopping, cooking), one action for each ingredient that was plated, one action for serving a plate and one action each for delivering dirty dishes, washing dishes or placing clean dishes on a table.
PCA Duration	Assessed how long, in seconds, that it takes to complete a single PCA
Chef Role Contribution (CRC)	The relative proportion of confederate PCAs subtracted from the relative proportion of participant PCAs. Positive CRCs indicate that the participant contributes more, negative CRCs that the confederate contributes more.

III. Procedure

Upon arrival in the CHOAPT testbed GRILL West laboratory, the participant was provided with consent information. Once verbal consent was obtained, participants were asked to complete a biographical data questionnaire which included items about prior video game and Overcooked 2! experience, personality and initial perceptions of team affect. The experimenter then provided a brief description of the game concepts and objectives, the game controls, and the condition in which they would be playing (human teammate vs. autonomous agent teammate). In the human co-located conditions, the experimenter served as the confederate and played physically in the same laboratory space (see Figure 2). In the human and autonomous player remote conditions, the confederate player, whom the participant did not meet, played in a separate room nearby. Participants would then play through the assigned levels and attempt to earn 3 stars on each before moving on to the next level. Video and audio was recorded throughout the game play session. At the end of about 35 minutes of gameplay, participants were asked to complete the affective team measures again. Participants were then provided with debriefing information and awarded course credit for their participation. The entire study took approximately 50 minutes to complete.

RESULTS

Due to the experimental design of the study not all participants completed six levels (see Table 2). This resulted in a substantial amount of missing data in the later levels of the game. Additionally, seven participants completed the human co-located condition, and only one participant each completed the human-remote and autonomous-remote condition. There was thus not enough data to compare the originally designed agent manipulation. A subset of the available dataset was therefore analyzed. From the sample of nine participants across six levels, eight participants were analyzed for the first three levels. For the communication data, only four participants were analyzed.

TABLE 2. NUMBER OF LEVEL ATTEMPTS AND TEAM COMPLETIONS

Level Name	**L1-2**	**L1-3**	**L2-1**	**L2-3**	**L2-4**	**L2-5**
Food Type / Difficulty	Sushi smpl	Sushi cmplx	Chicken River	Pasta cmplx	Burrito smpl	Burrito cmplx
Completions	9	9	8	7	6	5
Attempts	1	1.2	2.8	2.3	1	1.2

I. Team Perception

Team perceptions were generally high ($M = 5.69$, $SD = 1.28$). The subjective data was submitted to a one-way repeated measures MANOVA with Experience (Pre, Post) as the within-subjects variable and team trust, team affect, team cohesion, team collaboration, team effectiveness, and role clarity as the composite dependent measures. The MANOVA was not significant, $F(6, 3) = 1.19$, $p = 0.48$. Univariate analyses further revealed that only perceptions of role clarity was significantly higher after gameplay ($M = 5.89$, $SE = .30$) compared to before game play ($M = 5.02$, $SE = .39$), $F(1,8) = 6.36$, $p = 0.036$. This result makes intuitive sense because much of the game play in Overcooked is focused on division of responsibilities through clearly defined roles.

II. Game Experience

The usability ($M = 5.52$, $SD = .90$), enjoyment ($M = 5.67$, $SD = 1.03$) and social experience ($M = 5.61$, $SD = .70$) of Overcooked! 2 were all rated highly (scale 1-7). The participants' personality trait conscientiousness significantly and positively predicted enjoyment, $r = 0.67$, $p < 0.05$.

III. Game Score and Productive Chef Actions

The first step in understanding team performance in Overcooked!2 was to examine the relationship between game score and productive chef actions (PCAs) for the team, the confederate and the participant. The data was submitted to a one-way repeated measures MANOVA with Level (L1-2, L1-3, L2-1) as the within-subjects variable and game score, expertise score, PCAs, PCA duration, and chef role contribution together as the composite DV. The MANOVA was significant, $F(10, 22) = 34.80$, $p < 0.001$. Follow-up univariate analyses showed that game score decreased linearly across levels, $F(1,7) = 64.28$, $p < 0.001$ (see Figure 3). In similar fashion, team PCAs decreased linearly across this set of levels, $F(1,7) = 151.78$, $p < 0.001$. Indeed, game score and PCAs are highly positively correlated, $r = .84$, $p < 0.01$. Investigating team player PCAs further, a mixed repeated measures ANOVA with Player (Confederate, Participant) as the between-subjects variable and Level (L1-2, L1-3, L2-1) as the within-subjects variable revealed a significant interaction, $F(1.17, 16.36) = 7.90$, $p = 0.01$ (Greenhouse-Geisser adjustment). Confederates had more PCAs than participants in the simple L1-2 level, but this trend reversed for the third, and far more complex, L2-1 level. Overall, these results show that differences in level difficulty, and the unique contributions of the confederate and player actions to the team can be disentangled using a combination of measurement assessments.

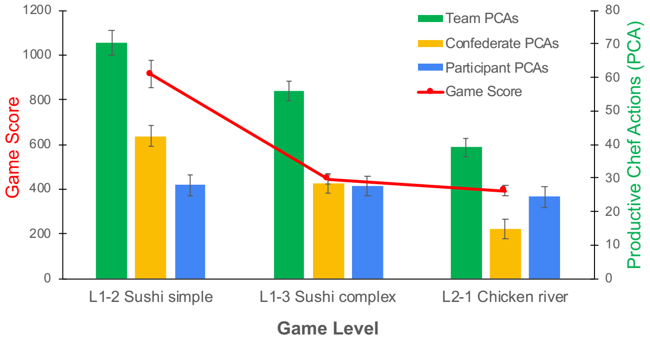

FIGURE 3. AXES INDICATE GAME SCORE (LEFT) AND PCAS (RIGHT)

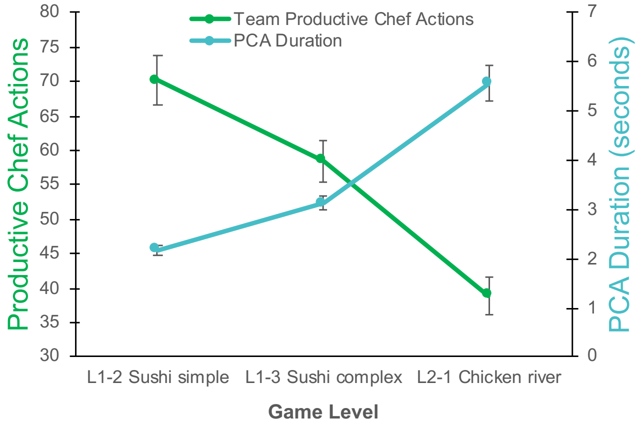

FIGURE 4. AXES INDICATE TEAM PCAS (LEFT) AND PCA DURATION (RIGHT)

IV. Productive Chef Actions and PCA Duration

Each level in Overcooked presents a unique challenge, thus it may be difficult to compare and classify behavior across levels. Each level also has its own time limit. We created the PCA duration measure to normalize the time, in seconds, that it takes to complete one single PCA. A follow-up univariate analysis demonstrated that PCA duration increases linearly across this set of levels, $F(1,7) = 158.08$, p < 0.001 (see Figure 4). Not surprisingly, PCA duration is also highly negatively correlated with Team PCAs, $r = -.95$, $p < 0.01$. As can been see in Figure 4, the time it takes to complete 1 single PCA increases as a function of the levels. Because it takes longer to complete 1 PCA, fewer team PCAs can be produced, which may lead to a lower game score. Indeed, PCA duration and game score are negatively correlated, $r = -.75$, $p < 0.01$. Increases in PCA duration are often an indication of the increased coordination that is required in a level due to environment layout, recipe complexity, and food order diversity (how many different recipes are ordered).

V. Game Score and Team Expertise Score

Another way to make levels more comparable might be to normalize the game score. Upon completion of a segment of the game, players are given an opportunity to "4-star" a level, which usually tests players to the maximum of their ability both individually and working together. Scores required to obtain Four-star are determined by the game based on the number of players playing the specific level. Figure 5 shows the game score of each level and the scores divided by the 4-star expertise score. A repeated measures ANOVA revealed a significant effect for Level on expertise score, $F(1.21, 14) = 9.42$, $p = 0.01$ (Greenhouse-Geisser adjustment).

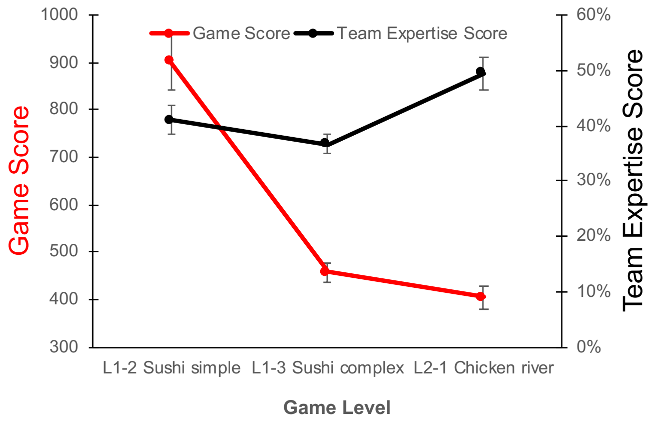

FIGURE 5. INDICATE GAME SCORE (LEFT) AND TSE (RIGHT)

Bonferroni corrected post-hoc comparisons showed that expertise score was higher in L2-1 ($M = .49$, $SE = .025$) compared to L1-3 ($M = .37$, $SE = .015$), $p < 0.01$. There was no difference between L1-2 ($M = .42$, $SE = .027$) and L1-3 (M = .37, SE = .015), $p = .24$). Game score shows the opposite trend and decreases across these levels (see Figure 5). Indeed, game score and expertise score do not significantly correlate, $r = .17$, $p > 0.05$. Expertise score thus tells a different story of gameplay compared to game score. The ability of players may get better over time and as they play more levels, but level difficulty at the same time goes up. Raw game score may therefore not always be the best indicator to capture this growing expertise. The normalized expertise score may be more useful in this regard.

V. Expertise and Chef Role Contribution

A measure was created to examine the unique chef role contribution (CRC) of each player. A repeated measures ANOVA revealed a linear increase in CRC across levels, $F(1,7) = 7.69$, $p < 0.05$. Participant contribution steadily increased as the team played more levels together. Team expertise score and CRC are also significantly positively correlated, $r = .49$, $p < 0.05$ (see Figure 6).

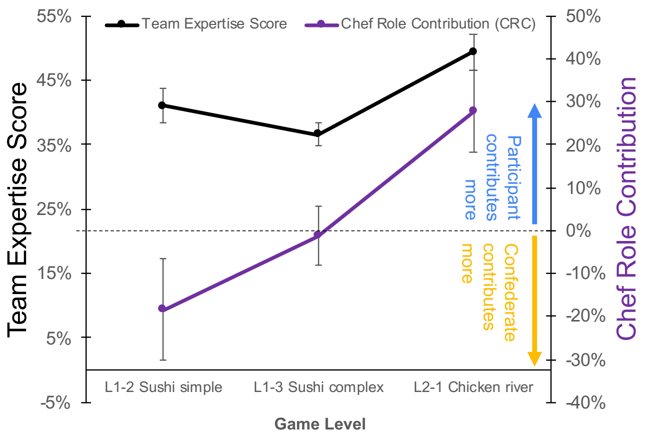

FIGURE 6. AXES INDICATE TSE (LEFT) AND CRC (RIGHT)

This indicates that, at least for this sample, a major contributor to an increase in team expertise is increases in participant contributions. The CRC may also be useful as a measure for the natural environmental asymmetry that occurs between levels. For example, in L2-1, the kitchen is split onto two rafts on a river physically separating the two players. In our experiment, the participant was typically assigned to the lower raft, which was equipped with most of the kitchen capability (e.g., cooking, plating, delivering). The participant therefore had to do the heavy lifting which caused an increase in PCAs and thus also an increase in the CRC. It is more difficult to separate player ability from kitchen asymmetry. The CRC may thus be an early indicator of a disparity of player ability or kitchen asymmetry.

TABLE 3. EXAMPLE COMMUNICATION VIGNETTE FOR LEVEL 1-3. BLUE INDICATES A PULL AND GREEN A PUSH FOR INFORMATION

Player	Communication Event
Participant	Dirty plates need to be washed? (Confederate confirms)
Confederate	The sink is over on your side
Confederate	I'm going to throw you some rice
Participant	Do I need to cook it?
Participant	How do I know when it's done? (Confederate clarifies)
Participant	Green bar, so this is done? (Confederate confirms)
Confederate	If you throw me a cucumber I can chop it

VI. Communication

Pushing information (i.e., providing information without a request) has been associated with improved team performance [3]. Push ratio and communication frequency were analyzed with a mixed repeated measures ANOVA with Agent (Confederate, Participant) as the between-subjects variable and Level (1-5) as the within-subjects variable. For push ratio, there was a main effect for Agent, $F(1,6) = 15.33$, $p < 0.01$. Confederates exhibited a higher push ratio ($M = .50$, *SE* = .05) compared to participants (*M* = .21, *SE* = .05). For communication frequency, there was a main effect for Level, *F*(4,24) = 4.14, *p* < 0.05. Post-hoc comparisons revealed that communication frequency was lower in Level 2-1 (*M* = 4.5, *SE* = .89) compared to both Level 2-3 (*M* = 9.63, *SE* = .72), *p* < 0.05, and Level 2-4 (*M* = 12.25, *SE* = 2.70), *p* < 0.05. Table 3 shows an example dialogue.

Discussion

The purpose of this research was to assess the viability of Overcooked! 2 as a research platform to assess human-autonomy teaming and validate behavioral measures that could be used as metrics of team process and performance. Unique measures derived from this platform such as productive chef actions, team expertise score and chef role contributions were all sensitive to differences in levels and between participant and confederate. Our findings demonstrate that we can derive meaningful team process, performance, and communication measures and that the interactions with this video game meet the requirements of psychological fidelity of human-autonomy teaming. We discuss several ideas for further research and measurement.

There are many avenues and research directions that could be pursued with the CHAOPT testbed. For example, *Autonomy Support Style* could be manipulated by varying the behavioral style of the autonomous player by executing either supportive, clumsy or self-centered play. Supportive play may increase trust, clumsy play may decrease competency-based trust and self-centered play may decrease integrity or benevolence-based trust. Team performance and affect may further be affected by *Autonomy Communication Mode*, a manipulation that compares Overcooked's pre-set emoji-like commands used for in-game communication with free-flowing audio communication. Another idea may be to manipulate *Team Training*. Human-autonomy teams that implement formal teamwork strategies, such as role-clarification exercises, may perform at a higher level than teams that do not [8]. Lastly, the CHAOPT testbed could investigate *AI Team Composition*. Previous research has indicated that the composition of human and AI agent teams may have differential effects on team performance and trust [10].

Measurement could further be improved by capturing and quantifying additional metrics in the game. For example, much of the game strategy and teamwork revolves around creating efficiencies to reduce the time it takes serve food order. For each level, a custom workflow rhythm between players must be developed to achieve high performance that includes strategies such as efficient kitchen navigation, accurate throwing, clever placement of ingredients and the flexible role switching towards the end of a game session. Unproductive chef actions may include trash dumps, making dubious food, and falling down chasms. Capturing these (in)efficiencies with additional measures will better characterize team expertise. Another way to improve measurement is by enhancing the precision of the PCA measure. The PCA in this paper did not include each individual action to prepare a food order which across levels can include picking up, throwing, cutting, and cooking ingredients. This also prevented us from capturing, in more detail, team supportive backup behaviors that can occur such as setting up or throwing an ingredient; a common occurrence. A revised PCA measure should incorporate this nuance. Lastly, level difficulty can be better characterized by quantifying navigation efficiency for kitchen layouts, recipe complexity (number of steps required to prepare each unique recipe) and food order diversity (how many different recipes appear in the food order cue). With these improvements, more precision in team process and performance measurement using Overcooked!2 can be achieved.

Conclusion

This paper has demonstrated the utility of using the video game Overcooked!2 for human-autonomy team performance research. The CHAOPT testbed is a rich platform for conducting human-autonomy teaming research.

Acknowledgments

This material is based upon work supported by the Air Force Office of Scientific Research under award number 16RT0881 and NSF Grant IIS-1909847.

References

[1] Goetz, A., Keebler, J. R., Phillips, E., Jentsch, F., & Hudson, I. (2012). Evaluation of COTS simulations for future HRI teams. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 2547-2551). Sage CA: Los Angeles, CA: SAGE Publications.

[2] Tossell, C., Donadio, B., Stewart, A., Tenhundfeld, N., Phillips, E., Driggs, J., ... & de Visser, E. (2019). Human Factors Capstone Research at the United States Air Force Academy. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 63, No. 1, pp. 498-502). Sage CA: Los Angeles, CA: SAGE Publications.

[3] McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human factors*, *60*(2), 262-273.

[4] Salas, E., Wilson, K. A., Lazzara, E. H., King, H. B., Augenstein, J. S., Robinson, D. W., & Birnbach, D. J. (2008). Simulation-based training for patient safety: 10 principles that matter. *Journal of Patient Safety*, *4*(1), 3-8.

[5] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ... & Lillicrap, T. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, *362*(6419), 1140-1144.

[6] Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., ... & Oh, J. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, *575*(7782), 350-354.

[7] Riek, L. D. (2012). Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, *1*(1), 119-136.

[8] Walliser, J.C., de Visser E. J., Wiese E., & Shaw, T. H, (2019) Team structure and team building improve human-machine teaming with autonomous agents. *Doctoral Dissertation,* George Mason University.

[9] Phan, M. H., Keebler, J. R., & Chaparro, B. S. (2016). The development and validation of the game user experience satisfaction scale (GUESS). *Human factors*, *58*(8), 1217-1247.

[10] Hertz, N., Shaw, T., de Visser, E. J., & Wiese, E. (2019). Mixing It Up: How Mixed Groups of Humans and Machines Modulate Conformity. *Journal of Cognitive Engineering and Decision Making*, *13*(4), 242-257.