ARL-TR-9359 ● DEC 2021

A Mobile Data Collection System for Studying Human Autonomy Teaming in Conjunction with Passive Context and Psychophysiological Sensing

by Torin Adamson, Weichen Wang, Yazied Hasan,
Andrew Campbell, Lydia Tapia, and Evan Carter

NOTICES

Disclaimers

ARL-TR-9359 • DEC 2021

A Mobile Data Collection System for Studying Human Autonomy Teaming in Conjunction with Passive Context and Psychophysiological Sensing

by Torin Adamson, Yazied Hasan, and Lydia Tapia
Department of Computer Science, University of New Mexico

Weichen Wang and Andrew Campbell
Department of Computer Science, Dartmouth College

Evan Carter
Human Research and Engineering Directorate, DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE	***Form Approved*** ***OMB No. 0704-0188***

1. REPORT DATE *(DD-MM-YYYY)* December 2021	**2. REPORT TYPE** Technical Report	**3. DATES COVERED (From - To)** 1 October 2020–30 September 2021

4. TITLE AND SUBTITLE A Mobile Data Collection System for Studying Human Autonomy Teaming in Conjunction with Passive Context and Psychophysiological Sensing	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Torin Adamson, Weichen Wang, Yazied Hasan, Andrew Campbell, Lydia Tapia, and Evan Carter	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) DEVCOM Army Research Laboratory ATTN: FCDD-RLH-FA 2800 Powder Mill Road, Adelphi, MD 20783-1138	**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-9359
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

It is expected that the ability for autonomous agents to adapt to their human teammates will be a critical development in building effective human-autonomy teams for the future. We argue that the current state of research in this area is limited by high experimenter and participant burden (e.g., traveling to a laboratory location; single, multi-hour experimental sessions; expensive, custom experimental set-ups). As an alternative, we develop a data collection system that makes use of a mobile videogame platform in conjunction with passive sensing of context and psychophysiology through the phone and a wearable device. This design dramatically reduces burden while maintaining flexible experimental design and multi-modal measurement. Here, we describe this novel system, as well as an illustrative research design for how it would be used to develop autonomous agents with adaptive capabilities. We also present a small set of pilot data validating several aspects of the system.

15. SUBJECT TERMS

human–autonomy teaming, mobile sensing, adaptive autonomy, mobile videogames, ambulatory assessment

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Evan C Carter
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	32	**19b. TELEPHONE NUMBER (Include area code)** (240) 478-9295

Standard Form 298 (Rev. 8/98)

Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

1. Introduction

The pace of advancements in artificial intelligence (AI) and machine learning being integrated into daily life indicate that future military teams will work closely with intelligent agents capable of making autonomous decisions. One potential benefit of this trend is that intelligent agents may allow for increases in performance through a capacity to adapt and personalize themselves to individual operators and situations. Modern doctrine relies on assumptions about uniform capabilities across operators—a kind of "one size fits all" approach—that, although it promotes robustness through interchangeability of team members, potentially limits above-average individuals. Thus, the promise of integrating adaptive intelligent agents into systems is that they can tailor themselves to a user and allow the individual to maximize his or her performance while maintaining robustness.

In addition to possible changes at the level of the individual, improvements in intelligent agents will likely have some effect on teams.[1] For example, technology such as night vision, although sophisticated, endows only an individual Soldier with greater capability without fundamentally shifting the way in which a team works together. In contrast, the introduction of intelligent agent technology holds the possibility of entirely shifting the team dynamic, just as would be the case when adding new team members and modifying team roles. As such, overall team performance will be a function of the degree to which the human and autonomous teammates can work well together, for example, adapting to one another's strengths and weaknesses. In contrast, if agents perform their roles in a rigid manner, it is likely that human teammates will experience their addition only as an increased burden.

For these reasons, the study and development of adaptive intelligent agents may yield considerable benefits for the Soldier of the future. One difficulty facing such work, however, is that approaches to developing adaptive agents can be notoriously data greedy. That is, training data must sufficiently cover the (potentially massive) variability represented by many people in many situations for one to develop robust algorithms on which agents can base their efforts at personalization. Depending on the complexity of the target task for which an agent is being developed, this can be prohibitively complicated, expensive, and time consuming. Moreover, if the signals that will be used to trigger adaptation are difficult to measure, such as a psychological construct captured by multiple variables, or occur at longer timescales, such as the accumulation of stress and fatigue, the task of data collection can become exponentially more difficult. For example, in the literature on the topic of this work—navigation and collision avoidance by one human and one autonomous driving assistant—data collection efforts involve cross-sectional measurements using highly customized, often expensive experimental set-ups.[2–4] Typically, both research participants and experimenters must travel to the study location and spend multiple hours on-site. As a result, the standard approach to investigating this topic incurs a burden that can quickly limit both the amount of data collected and the timescale over which it is collected.

Adamson et al.[5] proposed a potential solution to this problem by the use of a low-cost, easily customizable mobile videogame called Busy Beeway. The advantage of this approach is that by making use of increasingly ubiquitous smartphones, it dramatically lowers the burden of collecting data from many people over long periods of time. Additionally, to the extent that a game is enjoyable, burden on the participant is reduced by increasing an intrinsic motivation to play. Of course, there are some disadvantages to the use of mobile research games as compared to the usual approach described previously. First, the researcher loses a significant amount of experimental control. This problem can be at least partially mitigated through the sheer volume of data it is possible to collect, and its impact is further lessened through a hybrid approach in which data collected “in the wild” are combined with data collected under strict laboratory conditions.

Motivating this work is a second issue: mobile games only collect purely behavioral data whereas a full laboratory set-up can include the measurement of relevant psychophysiological signals. We report on efforts to combine Busy Beeway with an existing software application, StudentLife,[6,7] that has been used to passively and continuously measure research participants through multiple modalities (e.g., physical activity, social interaction) and over extended periods of time. Additionally, we report the novel integration of StudentLife with a smartwatch to collect psychophysiological data. In sum, this novel system has the potential to drive the development of adaptive intelligent agents by dramatically increasing the scope and scale of data collected on humans completing a military-relevant task alongside an autonomous teammate.

2. Method

2.1 Busy Beeway

Busy Beeway is a mobile videogame platform for studying human-autonomy teaming during navigation and collision-avoidance in stochastic, dynamic environments. As mentioned, it was designed to be an engaging, low-cost, low-risk way to collect large amounts of data.

2.1.1 Gameplay

In Busy Beeway, the research participant navigates a bee avatar (1 in Fig. 1) from subgoal to subgoal represented by flowers (2) and then to a portal as the final goal (3). The participant moves the bee avatar through a screen-based joystick (6), but if they remove their finger from the screen, an autonomous navigation algorithm will take control of the avatar. At all times, an arrow in front of the avatar indicates the direction of the nearest goal (8). The level ends upon reaching the final goal. Performance is measured by the time it takes to complete the level (4); however, there are stochastically moving obstacles drawn as wasps onscreen (5) that, if the bee avatar collides with, result in an immediate loss and a level restart. Each level can be attempted three times. The game therefore requires both safe and fast navigation, and because the obstacles change speed stochastically, there is no guaranteed algorithmic solution.

Fig. 1 Gameplay interface of Busy Beeway. The goal is to move the player character (1) to collect all of the subgoals (2) and reach the final goal (3). The current play time for the attempt is shown at (4) to score the game. A lower time is a better score. Colliding with any of the dynamic obstacles (5) will result in failure. To assume manual control of the character, the player may use the on screen joystick (6). The current state of control and the number of retries left are shown at the top right (7). An arrow will appear around the player indicating the direction to reach the next subgoal or final goal (8).

In a previous study, participants ($N = 32$) played the game for 15 min with access to an agent that failed approximately 45% of the time on the hardest level.[5] Participants were not given any instruction about relying on the agent, but tended to make use of it—59% used the navigation at least half of the time, suggesting that the game induces a mix of both manual and automatic play. The game was also engaging—81% of participants agreed with the statement “I would like to play more to contribute to human-automation collaboration research.” Furthermore, distinct high-performing strategies were observed. Among top-performing participants that were more successful than the agent alone, several relied on the agent almost exclusively (greater than 97% of the time) and one almost never (less than 3% of the time). These different playstyles indicate that adaptive agents will need to be capable of understanding and responding to individual differences in strategy, and being understood by the human operator through legible behavior if they are going to improve overall performance together.

2.1.2 Illustrative Experiment Setup

It is possible to design the adaptive capacity of the autonomous agent in a variety of ways. One approach relies on a so-called just-in-time adaptive intervention (JiTAI)[8] developed through what is known as a micro-randomized trial (MRT).[9] In an

MRT, participants are exposed many times to experimental manipulations while additional “context” features and performance metrics are simultaneously measured. Using these data, a mapping between context, experimental manipulation, and performance can be learned. A JiTAI would then use this mapping to pick a manipulation to implement for a given context to maximize performance.

As an example of building such a JiTAI-based adaptive agent within Busy Beeway, we have chosen to implement several updated features from the previous effort. These features facilitate long-term data collection, a requirement of MRT, the measurement of context, and the experimental manipulation of autonomous agents. In this conceptualization, the JiTAI functions by identifying the kind of autonomous agent that is optimal given the participant’s context.

To begin, we implemented different parameterizations of the autonomous navigation algorithm, based on artificial potential fields (APFs).[10] Four distinct agents are operationalized as four distinct APF parameterizations, each color-coded in the game (note the orange outline of the bee avatar in Fig. 1). These agent “personalities" serve as the experimental manipulation of an MRT in that participants are systematically exposed to agents that behave differently in different situations. In conjunction with some measurement of context, the resulting data set could be used to create an adaptive intervention whereby the agent changes its behavior (by changing the parameters of the APF it is using) in response to the context of the person it is paired with.

In addition to the distinct agents, a scheduling system has been put in place such that participants can only provide data at specific times. The goal is for the experimenter to have the ability to gain greater coverage of possible participant states by eliminating the participant’s ability to select when data are provided. For example, if a participant is especially sluggish in the morning, he or she may prefer to only provide data in the evening. As a result, it would be impossible to study the effect of such states on the participant’s behavior and how interventions might improve on the negative effects of fatigue. Notifications are dispatched, reminding participants to play when the next scheduled time is reached.

When a participant opens the game application, they are immediately directed to a splash screen (Fig. 2) on which an autonomy plays continuously in the background. If the participant opens the application outside of their allotted play period, they will not be able to advance past the splash screen but will be given the time of their next play period (1). If the participant opens the application during their assigned play period, they can select the “Participate" button to begin (2). Once a play session is complete, they can use the “Upload” button to upload their data to the server (see Section 2.1.3).

Fig. 2 Busy Beeway title splash screen. If opened outside of the next participation time, the next scheduled time is displayed (1). Otherwise, the session can be started by touching the participate button (2). Diagnostic information about the device’s performance and battery health can be seen across the top of the screen (3).

One of our goals is to ensure the game can be played on participant’s personal devices so that experiments can be done at scale without the limitation of needing to purchase and distribute smartphones. A difficulty inherent in this approach is that potential participants will have devices of widely varying quality, some of which will be insufficient to run the game application. We therefore implemented an easily accessible indicator (red text will appear at 3 in Fig. 2 if a problem exists) on the splash screen to test whether a given device is unable to achieve the required performance. Furthermore, a battery health indicator has been implemented to ensure that the device will support StudentLife. This appears at 3 in Fig. 2 and displays “cold”, “dead”, “good”, “overheat”, “over voltage”, and “unknown” matching the possible states defined by the Android application programming interface (API).

Non-compliance (i.e., not providing data) and drop-out are common problems in long-term data collection efforts.[11] Frequently, such issues are dealt with by providing participants with monetary compensation for providing data as an incentive. In the case of Busy Beeway, however, because the game can be completed entirely by the autonomous agent, it is possible for participants to still provide data without actually attending to the game (i.e., coasting). We have therefore implemented a "bonus task" that is easy to achieve if one pays attention. Specifically, out-of-place objects are placed quasi-randomly near the second or third subgoal in a level (4 in Fig. 2). At a pre-specified time point (e.g., the end of three levels), the application goes to a bonus quiz screen (Fig. 3) in which the participant is asked to remember the objects he or she was shown.

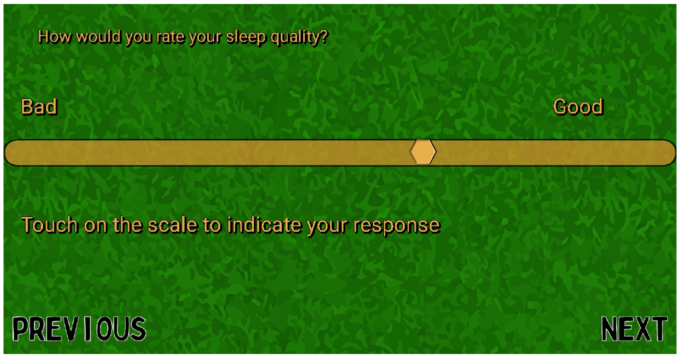
Bonus:

Which of these objects did you see while playing?

Pick the object

Fig. 3 Busy Beeway bonus quiz. To test if the participants were paying attention to the game, different objects appear around subgoals. At the end, the player must pick the three that had appeared from all possible objects.

As mentioned, the operationalization of an adaptive intelligent agent requires appropriate context measurement. Here, we define context broadly to include both the information available from passive sensing as done by StudentLife (see Section 2.1.3), as well as gameplay and information available via self-report by participants. To support this last point, we have implemented a variety of questionnaire item types directly into the Busy Beeway application (Fig. 4) including clock inputs, visual analog scales, and likert scales.



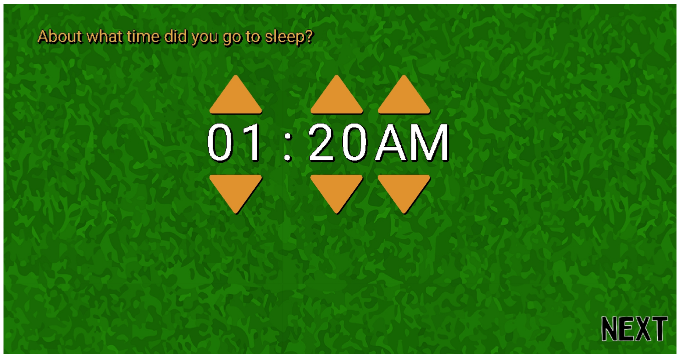
How would you rate your sleep quality?

Touch on the scale to indicate your response

(a) Analog Scale

How Upset do you feel?

(b) Likert Scale



About what time did you go to sleep?

(c) Time

Fig. 4 Examples of the different types of questionnaire prompts in Busy Beeway. Analog scale questions are from 0.0 to 1.0, five-point scale (Likert) questions are multiple choice, and time questions ask for a specific hour and minute of day.

2.1.3 Data Collection, Session Definition, and Deployment

Busy Beeway records raw screen input, the location of and control input to the bee avatar, information on all dynamic obstacles, and all other inputs from the user (e.g., non-control screen touches). As such, it is possible to fully reproduce the state of the game at any point as well as the participant's behavior. Data are recorded at 30 Hz to match the in-game frame rate, measured in frames per second (FPS). The game application currently only runs on Android operating systems. All of this study data is submitted by the participants while diagnostic data (crash events, etc.) are automatically collected. To monitor for problems, daily status reports that contain storage capacity information, the list of submitted participation sessions, and crash events are prepared and sent by the server to study personnel. A complete list of data collected can be seen in Table 1.

Table 1 Data collected from Busy Beeway gameplay. The sampling rate for this data is 30 FPS.

Type	Name	Description
All Objects	Position	Current position in game world (x,y)
	Velocity	Current velocity (x,y)
	Acceleration	Current acceleration (x,y)
	Direction	Angle of current heading, in degrees
Player	Control Input	Control input from player or AI (x,y)
	User control	If the player or AI was in control
	Touch input	Location being touched on screen (x,y)
	Touching	If the screen is being touched
	Gyroscope	Angular velocity along three axes (x,y,z) in radians/s
	Performance time	Nanoseconds taken to render frame

The actual parameters of a participation session, such as the form of agent, are defined in Extensible Markup Language (XML) files. All of the different conditions are stored on the server and a randomized schedule is generated per participant that combines every combination of play time (morning, day, or night) with each agent configuration so every possible condition is run once. Busy Beeway downloads the current day's experiment configuration from the server while the server tracks the actual day-by-day schedule. The XML files describe the kind of agent to be used and the associated parameters to set, the number of obstacles and the distribution of their stochastic dynamics, the number of subgoals, and the daily survey questions to ask. All of this can be updated as needed, even during a study, by modifying the XML files on the server. Notably, the parameters of each agent can be easily changed without requiring participants to download new versions of Busy Beeway. It is also easy to deploy alternative autonomous navigation algorithms such as A*[12] and Dynamic Risk Tolerance planning (DRT).[13]

Busy Beeway can be installed on participants' devices via Google Play through a closed beta invite. Future iterations of Busy Beeway can move toward open participation through Google Play seamlessly. This allows updates to be automatically delivered to participants with the option of checking the current version to require an update when the game app is started.

2.2 StudentLife

The StudentLife application, when installed on a smartphone, uses onboard sensors to collect GPS, accelerometer, microphone, application usage, and WiFi, cellular, and Bluetooth radio signals. Some of these data streams are directly interpretable (i.e., location, application usage, environmental light, and sound levels) and some are used to infer variables such as current activity (stationary, walking, and running), sleep duration, and sociability (i.e., the number and duration of conservation) via device-specific machine-learning pipelines.

Data collected using versions of StudentLife have been useful in uncovering results related to mental health. For example, Wang et al.[6] report that conversation frequency and duration were both related to lower levels of self-reported perceived stress, the degree to which a person feels situations are beyond their coping ability. In this same sample, self-reported depression was negatively correlated with several variables, including conversation frequency and duration, but also “co-locations”, or the number of Bluetooth devices in the same location as a subject. Interestingly, in another sample, data collected via a version of StudentLife were used to successfully train a model to predict the degree of resting state functional connectivity between the ventromedial prefrontal cortex and amygdala,[7] an index that has been shown to relate to negative outcomes such as anxiety[14] and depression.[15]

Data from StudentLife have also been used to predict achievement. Wang et al.[16] report successfully inferring the higher-order behaviors of studying and socializing from lower-order behavioral data collected by StudentLife. Both studying and socializing were then added to a model alongside self-report data and used to predict students’ GPAs. The resulting model suggested a kind of profile of students with better grades. They tended to study more, but they also tended to be less social in the evening, be more conscientious, and experience positive mood across the term and lower levels of stress as the term went on.

In this work, StudentLife collects participant data that is listed in Table 2 including the following:

- **Activity.** The activity that a participant is involved in is identified with the help of activity recognition API supported by Android phones. Activities include running, walking, in vehicle, on bike, on foot, tilting, still, and unknown. Data is collected continuously.

- **Phone Usage.** The mobile application records the number of phone locks and unlocks that the participants make. We derive (1) the number of phone locks and unlocks and (2) the average duration between phone locks and unlocks. The app also records the number and duration of incoming and outgoing phone calls and number of incoming and outgoing text messages (i.e., Short Message Service [SMS]). In both cases, the spoken content of conversations and written content of text exchanges is not recorded to preserve user privacy.
- **Mobility.** The embedded GPS, WiFi, and cellular tower location services provide an optimized location estimate. Locations are sampled every 10 min, with the consideration of both energy conservation and data quality. We use this information to derive (1) the number of unique locations visited, and (2) distance travelled.
- **Conversation duration.** The microphone is activated every 10 min to capture ambient sound. The software employs a voice-detection algorithm to recognize instances of human speech. To protect participant privacy, we use a speech-detection system that does not record raw audio on the device but instead destructively processes the data in real-time to extract, classify, and store features that are useful for inferring the presence of human speech but insufficient for reconstructing conversation content. The data collection software is active for the duration of human speech, logging start and end times.
- **Ambient light.** The app measures the ambient light conditions of the user's surrounding environment using light sensors on the phones.

Table 2 Types of data collected by StudentLife. Stress Level is based on HRV, and Body Battery is based off HRV, Stress, and Activity. * indicates metrics developed by Garmin. ** unspecified units proprietary to Garmin

(a) Collected Data

Name	Units
Activity	hours of activity (from category)
Voice inference	hours of noise or voice
Duration of conversations	hours
Phone usage	unlocks/h
Audio amplitude	**
Light amplitude	lx
Phone Battery	% of charge
Garmin Device Battery	% of charge
Heart rate (HR)	p/min
Heart rate variability (HRV)	ms
Oxygen blood saturation	% (SpO2)
Respiration rate	b/min
Stress level*	%
Body battery*	%

(b) Activity Cat.

Activity name
Running
Walking
In vehicle
On bike
On foot
Tilting
Unknown
Still

StudentLife has two features that are critical for use in mobile sensing research: the ability to provide privacy and confidentiality to participants and an online dashboard for researchers to monitor data quality. Three steps are used to preserve the privacy and confidentiality of participants (1) call and SMS logs are one-way hashed and therefore messages and phone numbers cannot be extracted from data; (2), both raw and inferred data are only temporarily stored on the local device before being uploaded using encrypted SSL connections and permanently removed when, for example, users have WiFi connection; (3) it is trivial to anonymize each participant and associate their data with a single study ID, as is best practice in human research. The online dashboard can be configured to output data streams relevant to a given study. In the current work, the dashboard displays one plot of each 24-h time series for each signal collected on each day for every participant. All signals mentioned previously are reported in a user-friendly dashboard.

The primary dashboard page displays a summary of data from all participants for the current day. Each row represents a single user, with cells representing the amount of hours collected by the app for various sensor/data categories (i.e., activity, speech inference, light, locations, Garmin data, etc.). Cells with insufficient data are colored red, so that a research manager will detect such missing data when monitoring the dashboard. A research manager can also choose different days and hone in on certain participants' data. They can determine whether or not to follow up with participants based on the details of missing data. (For example, data not collected during the night may be because the device was turned off. Alternatively, unavailable WiFi service could have interrupted data uploads to the server while the program continued to function.) To better interpret the reasons for missing data, the StudentLife app sends periodic small-size updates to a Slack messaging channel using the participant's mobile data plan if WiFi is unavailable. Such updates also include critical debugging information about the permissions the user has granted to the app as well as the services of sensors running on the app.

2.3 Garmin Vivoactive 4

The Garmin Vivoactive 4 is a smartwatch with the capability to collect data on heart rate and pulse oximetry. It was chosen to integrate with StudentLife because Vivoactive 4 was the only wearable at the time that offered: (1) a reasonable battery duration between full charges (4 to 5 days) and a reasonable charge time (charging from zero to full takes an hour or so); (2) a variety of sensors including an accelerometer, a barometric altimeter, a compass, GPS, a gyroscope, a heart-rate monitor, and a pulse oximeter that gauges oxygen saturation in the blood; (3) the ability to compute heart-rate variability (HRV) through beat-to-beat intervals (BBI) (BBI streaming requires a separate license) and consequently the ability to compute heart rate and respiration rate from HRV; (4) the ability to acquire useful health data such as sleep, stress level, body battery (a feature that uses a combination of HRV, stress, and activity to estimate a user's energy reserves throughout the day) and more using Garmin Health API.[17]

2.4 Integration

Busy Beeway and StudentLife operate independently of one another, yet the data can be combined for integrated analysis. Both associate an absolute time stamp in UNIX epoch format for each sample of data, allowing for event timelines to be constructed (an example of this can be seen in Fig. 5). The two systems could even be installed on separate systems, if need be. This effectively relies on the Network Time Protocol (NTP) to synchronize data and would only have an error of a few milliseconds.

After data is collected from both Busy Beeway gameplay sessions and StudentLife records, it can be synchronized using the millisecond time stamps. Figure 5 illustrates an example. A video demonstration of the integration in action is available. StudentLife, therefore, can provide context for Busy Beeway gameplay either in the very near-term (e.g., heart rate at a given time during gameplay) or at a more general level: for example, by looking at the entire time series of heart rate for a day, as well as other StudentLife data streams. Future research will need to be conducted to determine exactly what operationalizations of context are most relevant to team performance.


Time: 1630424318404ms (Absolute)
Heart Rate: 72.00 p/min
Heart Rate Variability: 731.00 ms
Oxygen Saturation: 94.00 %

Fig. 5 Rendering of Busy Beeway gameplay with selected data chosen from StudentLife. The time stamp value used to synchronize data is shown at the top left along with the current state of collected data for this frame of gameplay. The red dot indicates where on the screen the player is touching.

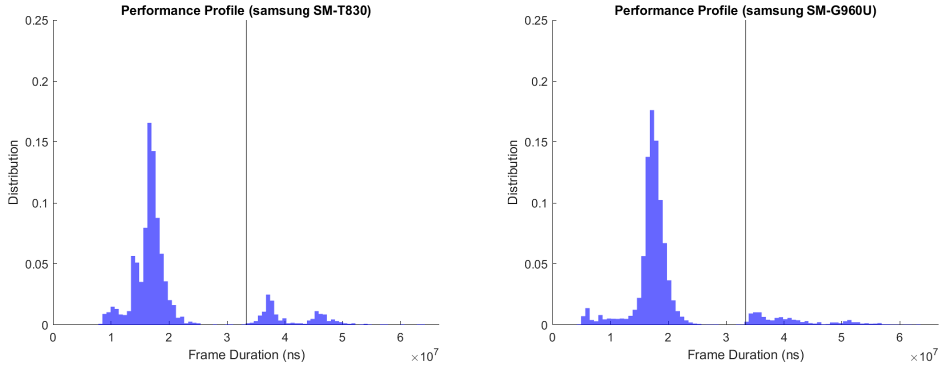
2.5 Pilot Tests

Busy Beeway was developed as a client server system with each participant's phone acting as the client. To evaluate the feasibility of the system as a crowdsource scale multi-study research platform, a series of pilot studies were conducted using the experimental design described previously. Testers in the research team acted as pilot participants. Device suitability tests, gameplay, and integration were assessed. Results are given in Section 3.

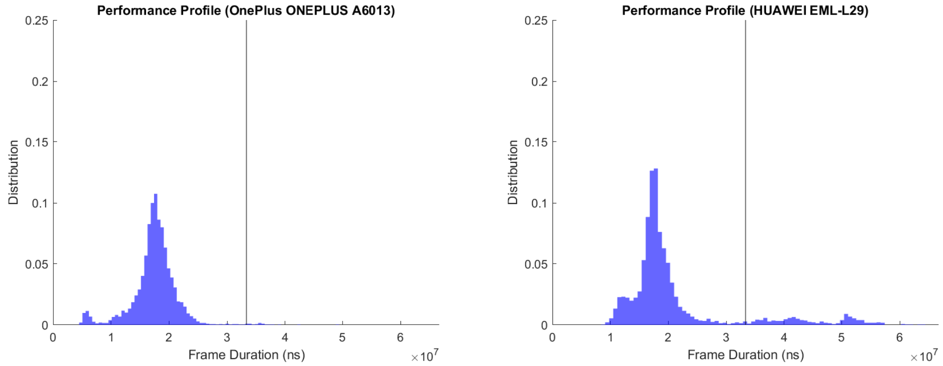
3. Results and Discussion

3.1 Device Suitability

Even though each particular participant's device can be automatically checked against study requirements, it is important to consider how the requirements limit the population of compatible devices as this would limit the amount of possible participants. Performance was profiled on a variety of consumer devices to estimate device compatibility by sampling frame rendering time, that is, the amount of time taken to prepare the next frame (in nanoseconds) as seen in Fig. 6. The target frame rate of the study directly affects the time limit allowed for rendering each frame, and therefore affects the amount of allowed devices. A frame rate of 30 is still acceptable for an action game[18] yet low enough that many devices can easily meet the requirement. Profiling is continued throughout the study to detect when devices no longer meet this requirement, usually due to physical damage or software changes.



(a) Samsung Tab S4 **(b) Samsung Galaxy S9**



(c) OnePlus 6T **(d) Huawei P20**

Fig. 6 Examples of performance profiles from devices running Busy Beeway. Performance is expressed in the amount of time taken to render each frame, in nanoseconds. The vertical line marks the time limit per frame to maintain the target frame rate (30 FPS). If most frames are finished in under this limit (to the left of the line in the graph), then the phone meets the performance requirements of the study.

3.2 Busy Beeway

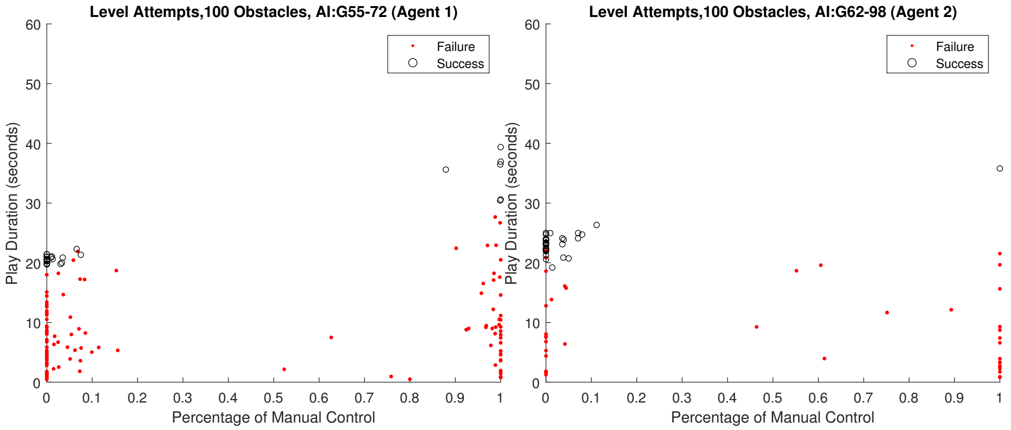
To test the effect of different AI behaviors on how players collaborate, AI agents are assigned each day from four possible configurations (referred to here as “Agent 1” through “Agent 4”). The parameters for each are listed in Table 3. Throughout the pilot study, gameplay attempts were recorded from daily testing by multiple testers (amounts shown in Table 4). Each attempt ends in either success (reaching the goal) or failure (collide with obstacle) with as few as 3 attempts in a day (all successes) or as many as 12 (all failures). For levels with 100 obstacles, the attempts are shown in Fig. 7 along with the amount of manual control and duration per attempt. Testers tended to let the agent run through the level by itself, with minimal intervention, or to assume near-complete control. While no example demonstrated that a tester exerting near complete control outperformed any of the four agents, the potential for one to do so can be seen in situations where the agent is more cautious (like Agent 3's larger detection radius) and the tester is willing to take risks for a lower completion time. These results are similar to the previous study,[5] which suggested that human-agent teaming with minimal human control interventions can produce reliable performance. Some notable exceptions previously seen were when a participant's skills were sufficient to exceed agent performance.

Table 3 Configuration of the four different kinds of APF based agents in this study. The detection radius is the distance at which the agent will react to obstacles, and the repulsion factor balances avoiding obstacles against traveling towards the next goal.

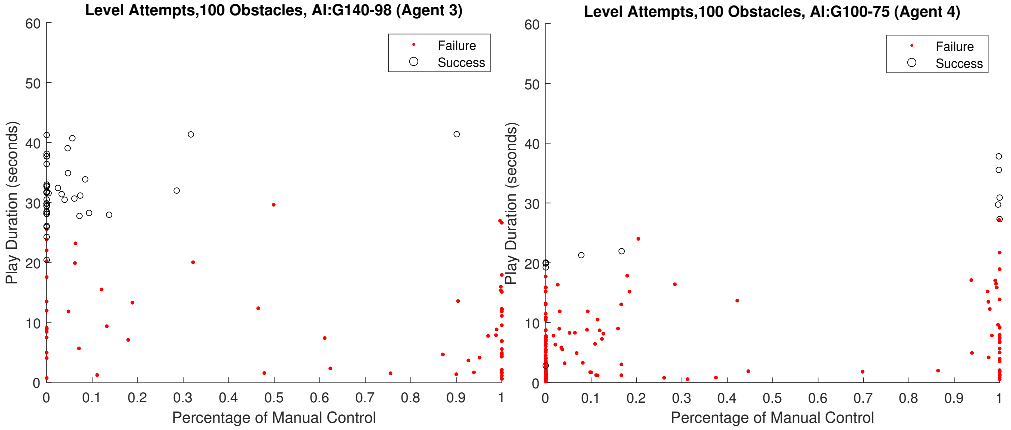
Name	Color	Detection radius	Repulsion factor
Agent 1	Orange	55	0.72
Agent 2	Light blue	62	0.98
Agent 3	Dark blue	140	0.98
Agent 4	Purple	100	0.75

Table 4 Amount of gameplay attempts collected per agent configuration. Each level has a different number of dynamic obstacles in the environment.

AI	Obstacles	Attempts	Successes	Success rate
Agent 1	50	99	49	49.5 %
	100	149	26	17.4 %
	200	166	4	2.4 %
Agent 2	50	61	39	63.9 %
	100	76	35	46.1 %
	200	123	14	11.4 %
Agent 3	50	56	47	83.9 %
	100	104	40	38.5 %
	200	142	10	7.0 %
Agent 4	50	129	38	29.5 %
	100	167	12	7.2 %
	200	165	0	0.0 %



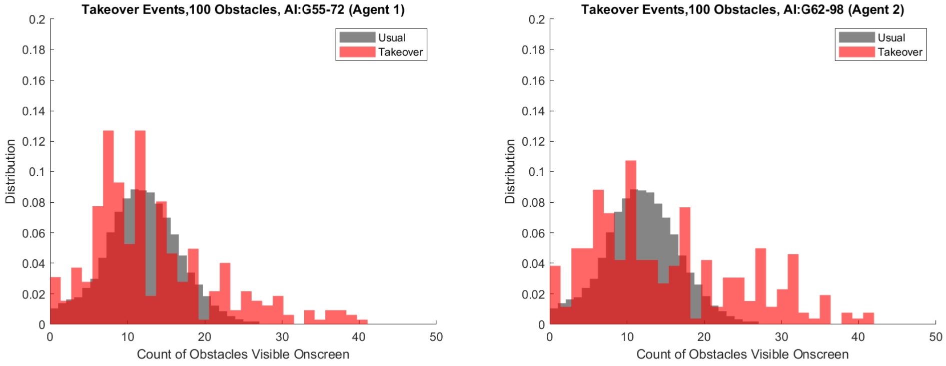
(a) Agent 1 **(b) Agent 2**



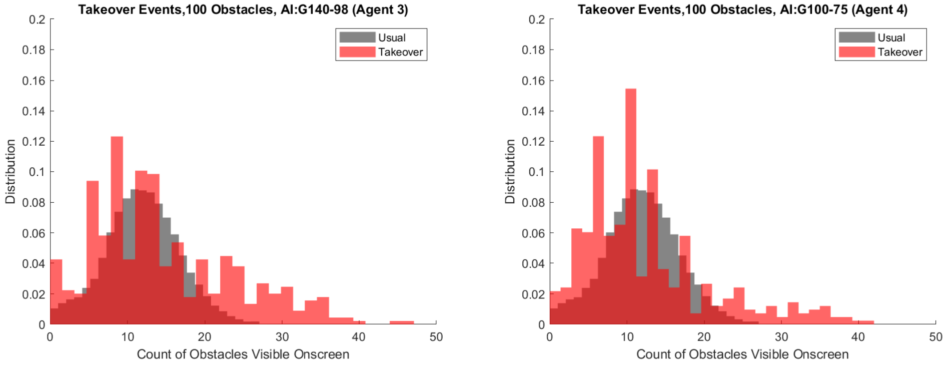
(c) Agent 3 **(d) Agent 4**

Fig. 7 Attempts by agent in an environment with 100 obstacles. The amount of manual control exercised along with the total gameplay duration is shown for each success and failure (collision).

Past studies of Busy Beeway investigated the self-reported strategy of taking control when many obstacles were visible onscreen.[5] Unlike the prior study, the testers did not demonstrate a pronounced difference in onscreen obstacle counts during takeover events from distributions normally encountered during play (see Fig. 8). We should note that this iteration of Busy Beeway renders more of the environment on the screen than the prior study, thus raising the overall numbers of on-screen obstacles from prior studies. The increased range of vision could also affect the behaviors of future participants.



(a) Agent 1 **(b) Agent 2**



(c) Agent 3 **(d) Agent 4**

Fig. 8 Count of obstacles on the screen during gameplay in an environment with 100 obstacles are shown in gray as "usual". Counts of obstacles for each takeover event (tester assumes control of the character) are shown in red.

4. Conclusions

Typical research on human-autonomy teaming makes use of highly controlled laboratory environments in which participants interact with immersive, expensive, custom setups. This approach creates a potentially large burden on both participants and experimenters, one which ultimately limits the amount of data collected. We contend that this limitation stands squarely in the way of progress on the development of autonomous agents with the capacity to adapt to their human counterparts. Therefore, we developed a data collection system that combines a mobile research videogame with continuous and passive sensing of context and psychophysiology.

As described previously, this novel system can be customized for use in long-term longitudinal experiments, or MRTs, that produce data on which adaptive capabilities for agents within the game can be based-for example, experimental manipulations of agent characteristics, multi-modal, continuous context measurement, a scheduling system, and auxiliary incentive tasks to maximize attention. Once adaptive agents are developed, they can be easily implemented into the game and additional experiments can be run to assess how human-autonomy teaming changes over longer timescales (e.g., due to learning effects or changes in factors such as sleep and stress), a task that would be highly costly under the standard approach to research in this area.

Ideally, we envision this system working in concert with experiments conducted through the typical in-laboratory method. This combination, with each approach shoring up the weaknesses of the other, strikes us as especially promising for the advancement of adaptive technology, as well as our understanding of how humans work with and respond to such technology.

5. References

1. DeCostanza A, Marathe A, Bohannon A, Evans A, Palazzolo E, Metcalfe J, McDowell K. Interactive paper: Enhancing human-agent teaming with individualized, adaptive technologies: A discussion of critical scientific questions. Army Research Laboratory (US); 2018 May. Report No.: ARL-TR-8359.

2. de Waard D, van der Hulst M, Hoedemaeker M, Brookhuis KA. Driver behavior in an emergency situation in the automated highway system. Trans. Hum. Fact. 1999;1(1):67–82.

3. Xu A, Dudek G. Maintaining efficient collaboration with trust-seeking robots. In: Proc. IEEE Intl. Conf. Intel. Rob. Syst. (IROS); p. 3312–3319.

4. Johns M, Mok B, Sirkin DM, Gowda NM, Smith CA, Talamonti Jr WJ, Ju W. Exploring shared control in automated driving. In: HRI ’16: The Eleventh ACM/IEEE International Conference on Human Robert Interaction; 2016 Mar 7–10; Christchurch, New Zealand. IEEE Press; c2016. p. 91–98.

5. Adamson T, Oishi M, Chiang HTL, Tapia L. Busy beeway: A game for testing human-automation collaboration for navigation. In: Proc. Intl. Conf. Motion Games (MIG); 2017 Nov; Barcelona, Spain. ACM; c2017. p. 9:1–9:6.

6. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, Zhou X, Ben-Zeev D, Campbell AT. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: UbiComp ’14: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2014 Sep 13–17; Seattle, WA. ACM; c2014. p. 3–14.

7. Obuchi M, Huckins JF, Wang W, daSilva A, Rogers C, Murphy E, Hedlund E, Holtzheimer P, Mirjafari S, Campbell A. Predicting brain functional connectivity using mobile sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2020;4(1):1–22.

8. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, Murphy SA. Just-in-time adaptive interventions (JiTAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med. 2018;52(6):446–462.

9. Walton AE, Collins LM, Klasnja P, Nahum-Shani I, Rabbi M, Walton MA, Murphy SA. The micro-randomized trial for developing digital interventions: Experimental design considerations. arXiv preprint arXiv:2005.05880. 2020;.

10. Khatib O. Real-time obstacle avoidance for manipulators and mobile robots. In: Proceedings of 1985 IEEE International Conference on Robotics and Automation; Vol. 5; 1986 Mar 25–28; St Louis, MO. IEEE; c1986. p. 90–98.

11. Eveleigh A, Jennett C, Blandford A, Brohan P, Cox AL. Designing for dabblers and deterring drop-outs in citizen science. In: Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI); 2014 Apr; Toronto, Ontario, Canada. p. 2985–2994.

12. Hart PE, Nilsson NJ, Raphael B. A formal basis for the heuristic determination of minimum cost paths. IEEE Trans. Sys. Sci. Cyber. 1968;4(2):100-107.

13. Chiang HT, HomChaudhuri B, Vinod AP, Oishi M, Tapia L. Dynamic risk tolerance: Motion planning by balancing short-term and long-term stochastic dynamic predictions. In: Proc. IEEE Intl. Conf. Rob. Auto. (ICRA); 2017 May; Singapore. p. 3762–3769.

14. Kim MJ, Loucks RA, Palmer AL, Brown AC, Solomon KM, Marchante AN, Whalen PJ. The structural and functional connectivity of the amygdala: from normal emotion to pathological anxiety. Behav Brain Res. 2011;223(2):403–410.

15. Connolly CG, Ho TC, Blom EH, LeWinn KZ, Sacchet MD, Tymofiyeva O, Simmons AN, Yang TT. Resting-state functional connectivity of the amygdala and longitudinal changes in depression severity in adolescent depression. J Affect Disord. 2017;207:86–94.

16. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, Zhou X, Ben-Zeev D, Campbell AT. StudentLife: Using smartphones to assess mental health and academic performance of college students. In: Mobile health; Springer; c2017; p. 7–33.

17. Garmin Health API. https://developer.garmin.com/gc-developer-program/health-api/; [accessed 2021 Sep 29].

18. Claypool KT, Claypool M. On frame rate and player performance in first person shooter games. Multimedia Systems. 2007;13(1):3–17.

List of Symbols, Abbreviations, and Acronyms

AI	artificial intelligence
APF	artificial potential Field
API	application programming interface
app	application
BBI	beat-to-beat intervals
DRT	Dynamic Risk Tolerance
FPS	frames per second
ID	identification
GPA	grade point average
GPS	global positioning system
HRV	heart-rate variability
JiTAI	Just-in-Time Adaptive Intervention
MRT	micro-randomized trial
NTP	Network Time Protocol
SMS	Short Message Service
SSL	Secure Sockets Layer
XML	Extensible Markup Language

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 DEVCOM ARL
(PDF) FCDD RLD DCI
TECH LIB

1 DEVCOM ARL
(PDF) FCDD RLH FA
E CARTER