2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Socially Assistive Robots at Work: Making Break-Taking Interventions More Pleasant, Enjoyable, and Engaging

Brian J. Zhang[1], Ryan Quick[1], Ameer Helmi[1], Naomi T. Fitter[1]

***Abstract*— More than ever, people spend the workday seated in front of a computer, which contributes to health issues caused by excess sedentary behavior. While breaking up long periods of sitting can alleviate these issues, no scalable interventions have had long-term success in motivating activity breaks at work. We believe that socially assistive robotics (SAR), which combines the scalability of e-health interventions with the motivational social ability of a companion or coach, may offer a solution for changing sedentary habits. To begin this work, we designed a SAR system and conducted a within-subjects study with N = 19 participants to compare their experiences taking breaks using the SAR system versus an alarm-like device for one day each in participants' normal workplaces. Results indicate that both systems had similar effects on sedentary behavior, but the SAR system led to greater feelings of pleasure, enjoyment, and engagement. Interviews yielded design recommendations for future systems. We find that SAR systems hold promise for further investigations of aiding healthy habit formation in work settings.**

I. Introduction

As computer use becomes more central to work in many fields, workers face increased risks of health challenges including heart disease [1], diabetes [2], and eyestrain [3], due to prolonged periods of sitting and looking at a screen. These individuals can benefit from taking breaks [4], standing up [5], and moving around [6], but there is no universally effective method for encouraging these behaviors in the workplace.

Previous work on reducing workplace sedentary behavior has shown that relatively simple e-health interventions, such as alarm-style reminders from phones, can significantly improve activity levels [5]–[7]. However, few interventions have shown sustained success, and the most successful techniques often incorporate expensive or non-scalable methods, such as human coaching and support [8], [9].

Socially assistive robotics (SAR) may offer a potentially groundbreaking solution to this problem. Past work suggests that prompts from physically embodied robots can encourage more cooperation over longer periods of use than onscreen prompts or other non-embodied methods [10], [11]. As the problem of workplace sedentary behavior grows increasingly widespread, SAR systems may provide more scalable, sustained, and motivational interventions when compared to current state-of-the-art solutions. Accordingly, we seek to understand if a social and physically embodied SAR system encourages healthy workplace behaviors more effectively than non-embodied methods.

[1] Collaborative Robotics and Intelligent Systems (CoRIS) Institute, Oregon State University, Corvallis, OR 97331, USA.
{zhangbr, quickr, helmia, fittern}@oregonstate.edu

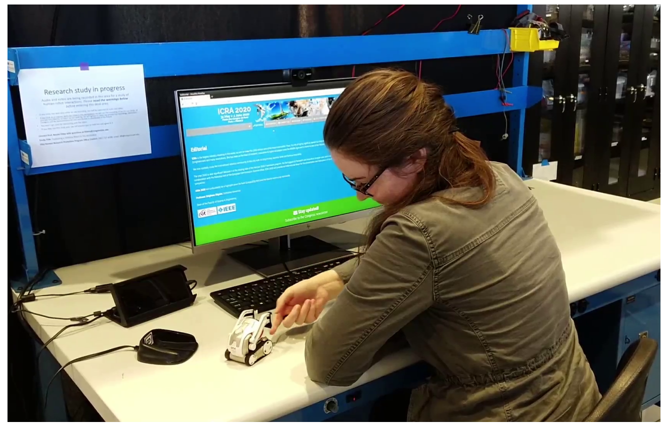

Fig. 1. Interaction between a computer user and our robotic system. The robot prompts the participant to stand through behaviors that combined driving motions, head motions, forklift motions, and facial expressions.

As the first step in investigating the suitability of SAR systems in the workplace, we aim to *identify advantages, disadvantages, and design constraints for robot-delivered break prompts compared to existing break-prompting methods.* We will incorporate the findings presented here into continuing work to confirm these findings and refine our SAR system.

With motivation from the related work detailed in Section II, this paper presents two break-prompting intervention tools: a *buzzer-based system* and a *robot-based system* (as seen in Fig. 1). We implemented the system hardware as described in Section III-B, and we evaluated both systems with recorded data, semi-structured interviews, and surveys, gathered during the within-subjects study outlined in Section III. The results in Section IV show differences between the studied systems, and the discussion in Section V may help inform other researchers developing SAR systems for encouraging healthier workplace behaviors.

II. Related Work

We reviewed literature and identified break-prompting interventions to reduce workplace sedentary behavior as a potential use-case for socially assistive robots.

A. Benefits of Activity Breaks

Sedentary behaviors, such as sitting, are on the rise in the modern workforce [1], [12], [13]. Past studies have shown that prolonged periods of sitting reduce metabolic health, increase cardiovascular morbidity and mortality, and increase the probability of diseases such as type 2 diabetes, certain cancers, and adolescent obesity [2], [14], [15]. Breaking up sedentary behavior with physical activity (e.g., standing, walking, or exercise) leads to significant benefits, including lowering mean arterial pressure [6] and lowering glucose and insulin levels after eating [5], [16], [17]. Our study accordingly focuses on a system for *punctuating periods of continual sitting with standing breaks*.

B. Reminder Technologies

Recent interventions for breaking up periods of sedentary behavior have involved reminder methods ranging from computer-based prompts to encouragement from supervisors [18]. Certain past interventions have succeeded in reducing sedentary behavior during studied periods, but longer-term benefits and break-taking adherence require additional investigation [19]. Modern phone and computer applications can facilitate the detection of periods of sedentary behavior [20], [21] and delivery of prompts at appropriate times to support the above efforts.

Specific intervention examples include phone applications that successfully encouraged increased physical activity levels [22] and reduced sedentary time [7] and a computer application that improved cardiovascular health [6]. Another computer application with prompts for microbreaks and stretch breaks increased computer worker productivity, evaluated through keyboard and mouse recordings [4]. Many free applications are available for turning phones or computers into automatic break reminders (e.g., [23], [24]). However, computer application prompts are often ignored, and even successful past studies have encountered break-taking noncompliance levels over 50% [25]. Our robotic system may be able to encourage *more cooperation with break prompts and better long-term health gains* than past work; compared to non-social or software-based prompting approaches, social and physically embodied robots are more likely to be viewed as a motivational companion.

C. Socially Assistive Robotics (SAR)

Socially assistive robotics (SAR) leverages the physical embodiment and social capabilities of robots to aid people in scenarios from physical therapy to healthy eating [26]. In previous work, SAR systems have successfully encouraged adults during post-stroke physical therapy interventions [27], led older adults through exercise routines [28], and promoted social skill practice by children with developmental delays [29]. These attainments are possible in part because of the embodiment of SAR systems; past studies have demonstrated that physical robots can encourage more cooperation, more positive feelings, and more feelings of bonding than onscreen prompts or agents can [10], [11]. Generally, SAR systems have considerable potential for cost-effective, scalable, long-term, and motivational interventions [30].

The closest preceding SAR research to our project is a prior study of Koosh ball-like robot prototypes that swayed as a break prompt for office workers [31]. This work focused mostly on the design principles for robots in everyday environments, rather than achieving better health practices in the workplace. Among the small user group in this prior study, participants took breaks 41% of the time when prompted by a non-social system and 65% of the time when prompted by a social system. Our study builds upon this prior SAR work, seeking to replicate the *successful encouragement* seen in past physical activity studies, deploy a *relatively complex SAR platform*, and run investigations with a *large enough sample size* to identify statistically significant differences.

III. Methods

We conducted a within-subjects study to compare the effects of a buzzer system and a robotic system for prompting periodic standing breaks. Participants used both following systems for one day each in their regular workplace. All study procedures were approved by the Oregon State University Institutional Review Board under protocol #IRB-2019-0067.

A. Hypotheses

Our two main hypotheses were based on the positive perceptions of SAR systems in [11] and the increased motivation from a physically embodied robot in [10]:

H1: The robotic system will lead to more pleasant and engaging interactions than the buzzer system.

H2: The robotic system will perform better than the buzzer system in encouraging standing breaks.

B. System Hardware

We created two systems with shared computing and sensing hardware but different interactive devices, as shown in Fig. 2. A Raspberry Pi 3 B+ running Ubuntu Mate 18.04 acted as the primary computer. Project-specific software was written in Python 3.5 or the Arduino language, and the ROS Melodic platform was used to easily interface with sensors.

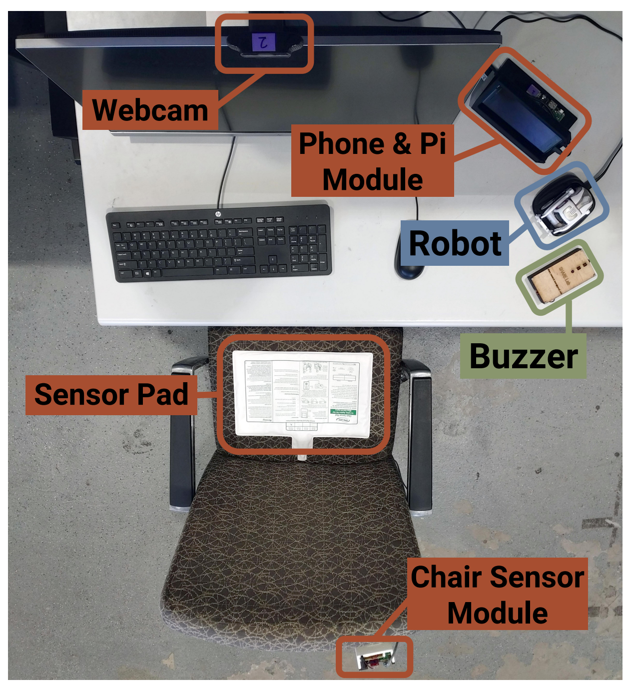

Fig. 2. Overhead view of both systems at a mock student desk. During deployments, only one of the two interactive devices (i.e., buzzer or robot) would be present. The remaining elements were present in both systems.

The Raspberry Pi computer connected to several peripheral devices: a Nokia 6 Android phone, a SparkFun Pro nRF52840 Mini microcontroller connected to a seat sensor, and a USB webcam. The Android phone communicated with the computer through a USB-based Android Debug Bridge in the robotic system, and was present as a visual placeholder in the buzzer system. For both systems, the phone screen remained unlit. The microcontroller in the chair sensor module sent data from a commercial seat sensor pad to the computer over a Python-controlled `gatttool` Bluetooth connection. The computer used `ffmpeg` with options `-c:v libx264 -preset ultrafast` to record audio and video of participant interactions via the webcam.

C. Study Design

The study had two conditions:

- **Buzzer System** (*control condition*): A custom buzzer device delivered break prompts. This wireless, alarm-like device contained a SparkFun Pro nRF52840 Mini wireless microcontroller, accelerometer, shaftless vibration motor, and blue LED. The system communicated with the computer using Bluetooth. When prompting, the buzzer would vibrate and flash its LED in a 1.25 Hz cycle for up to 60 seconds.
- **Robotic System** (*experimental condition*): An Anki Cozmo robot delivered break prompts. Cozmo is a small mobile robot with a programmable OLED face-like screen, movable head, and forklift-like mechanisms. The Cozmo system offers a suite of social behaviors in a small and relatively inexpensive package. We created Cozmo behaviors using the Cozmo SDK and controlled the robot from the computer through the phone, which ran the Cozmo application in SDK Mode. Prompt behaviors were selected randomly from a predetermined set and ran for 90 to 110 seconds. An example of these behaviors is demonstrated in the video included with this paper.

In both conditions, the device delivered a break prompt after every 30 minutes of continuous sitting and could be delayed for 10 minutes by being flipped over.

D. Participants

19 technical students enrolled in and completed the study. Participants were adults between 19 and 30 years of age ($M = 24.5$, $SD = 3.1$), with 13 male and 6 female participants. On 7-point scales, participants self-reported high prior experience with robotics ($M = 5.4$, $SD = 1.3$) and low prior experience with Cozmo ($M = 2.4$, $SD = 1.0$). Each participant received US$25 after completing the study.

E. Procedure

To begin, a research assistant explained that the study involved two break-prompting systems. The participant gave informed consent, saw a brief demonstration of both systems in order of assigned use, and completed a pre-study survey.

Next, the participant used each of our break-prompting systems for one day. Trial order was counterbalanced and randomly assigned across participants. The research assistant installed the first designated system in the participant's regular workplace and explained the system's use.

The participant would then use the system for an agreed-upon period of at least 3 hours. At the end of the day, the research assistant collected the system, and the participant completed a post-day survey. The following day, the research assistant would deploy and explain the other break-prompting system for the participant to use for the day. To conclude the second day, the participant completed the post-day survey, a post-study survey, and a semi-structured interview.

F. Measurement

Participants completed three types of surveys during the study. All survey question responses, other than two free response questions, used 7-point Likert scales.

- The pre-study survey captured participants' preconceptions about robots using a validated questionnaire based on the Unified Theory of Acceptance and Use of Technology (UTAUT) [32], and included questions about participants' robotics experience and health goals.
- The post-day survey, taken after both days of the study, captured participants' experiences during one study day using questions adapted from the Self-Assessment Manikin (SAM) [33] and the NASA Task Load Index (TLX) [34], and included questions for free comment.
- The post-study survey included all UTAUT-based questions, personality questions, and demographic questions.

At the end of the study, participants completed a semi-structured interview, which provided qualitative data on participant experiences and design requirements.

During the study, we gathered data through system logs and audiovisual recordings. Specifically, the system recorded when the participant sat, the timing of break prompts, the timing and type of participant responses to the prompts, and audio and video before, during, and after each break prompt. A Fitbit Inspire HR also tracked the heart rate, step count, and physical activity level of participants. We parsed this data to determine total participant sitting time, time from break prompts to standing behaviors, and prompt success.

To compare behavioral and survey data, we extracted emotion levels from facial expressions using OpenFace 2.2.0 on video recordings of each interaction. This software identified facial action unit (FAC) intensity (objective measures of facial muscle activations that correspond with emotion [35]) to evaluate participant *happiness*, *sadness*, and *anger* levels [36]. For instance, observed cheek raising and lip corner pulling contribute to the detected *happiness* level. We excluded frames with facial tracking confidence below 0.7 as recommended by the software author [37]. We also excluded videos in which 95% or more of the analyzed frames were below the confidence threshold. OpenFace analyzed video starting from 15 seconds before the device prompt through the duration of the interaction to find the change between average pre-prompt and post-prompt happiness, sadness, and anger for each participant. Lastly, we determined change in gaze variability using the same OpenFace data by determining

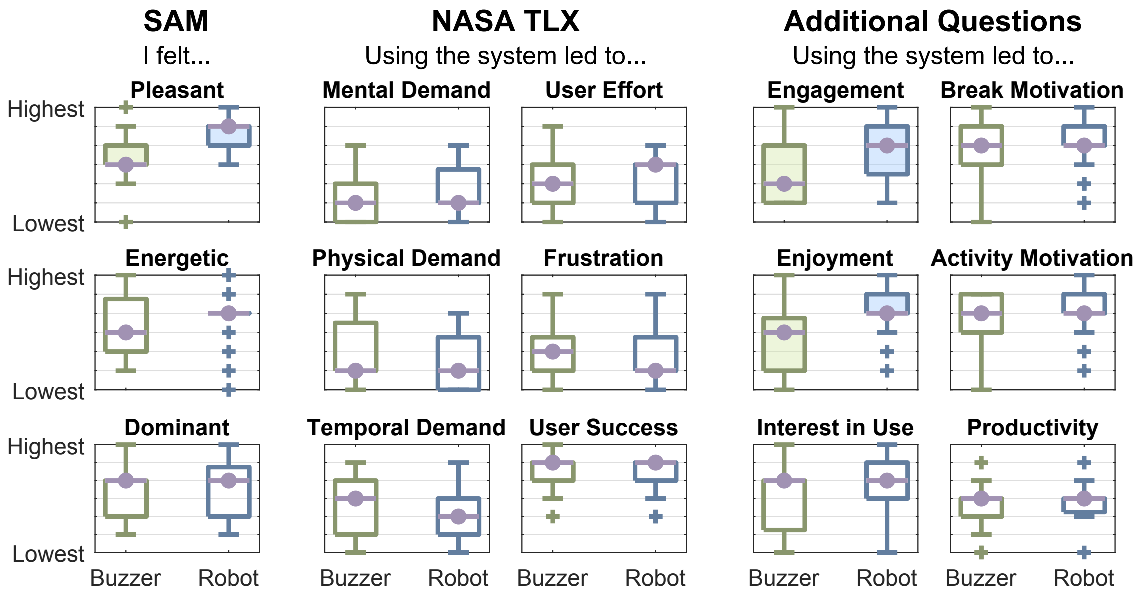

Fig. 3. Results of the post-day survey. Filled-in boxplots indicate significant differences. Purple lines with circles represent the median. Boxes range from the 25th to the 75th percentiles, and whiskers range up to 1.5 times the interquartile range, representing all non-outlier data. "+" marks indicate outliers.

the difference in the mean distance from the gaze centroid between pre-prompt and post-prompt excerpts.

G. Analysis

We analyzed survey and behavioral data using repeated-measures analysis of variance (rANOVA) and analysis of covariance (ANCOVA) tests with an $\alpha = 0.05$ significance level. We report effect size using η^2, where $\eta^2 = 0.010$ is small, $\eta^2 = 0.059$ is medium, and $\eta^2 = 0.138$ is large [38]. Regression analyses, also with $\alpha = 0.05$, helped us to identify meaningful covariates for ANCOVAs.

Similarly to [15], which considers interview data from an open-world study of a hospital robot, we used open coding techniques to identify and assess important topics in the interview data.

IV. Results

A. Survey Responses

Responses to post-day Likert scale survey questions appear in Fig. 3. After using the robotic system, participants' responses were significantly higher for *happiness* ($p = 0.007$, $F(1, 37) = 9.09$, $\eta^2 = 0.222$), *engagement* of interactions ($p = 0.029$, $F(1, 37) = 5.66$, $\eta^2 = 0.109$), and *enjoyment* of interactions ($p = 0.001$, $F(1, 37) = 14.57$, $\eta^2 = 0.274$).

There were no significant differences between the pre- and post-study survey responses, but preconception and demographic data from these questionnaires yielded useful covariates for understanding participant experiences. For covariate and correlation analysis, UTAUT questions were combined into their scale factors [32].

Correlation analysis showed that the more participants initially *accepted the robot* and *expected little effort* interacting with the robot, the less *frustrated* they reported being while using the studied systems (*acceptance*: Pearson's $r = 0.429$, $p = 0.007$; *effort expectations*: Pearson's $r = 0.321$, $p = 0.049$). *Expecting little effort* interacting with the robot also correlated with feelings of *energy level* in interactions (Pearson's $r = 0.336$, $p = 0.039$). Higher *expectations of performance* for the robot and *social reciprocity* with the robot were negatively correlated with self-reported break-taking *performance* (*expected performance*: Pearson's $r = 0.433$, $p = 0.007$; *social reciprocity*: Pearson's $r = 0.368$, $p = 0.023$). Because of these notable correlations, we chose to use these fours items as covariates in subsequent analysis.

ANCOVAs with the independent variable of system type and each covariate confirmed these correlation effects with the exception of the correlation of *effort expectations* with *energy level* of interactions. In each case, the effect of system type on *happiness*, *engagement*, *enjoyment* remained significant, and no covariates correlated significantly with these measures. Thus, we excluded them from the following results. Lastly, in all cases where the covariate had a significant effect, the system type did not have a significant effect.

Participants that *accepted the robot* more reported less *frustration* after system use ($p = 0.007$, $F(1, 37) = 8.21$, $\eta^2 = 0.135$); participants that *expected little effort* interacting with the robot reported greater *energy level* of interactions ($p = 0.035$, $F(1, 37) = 4.83$, $\eta^2 = 0.113$); participants that *expected higher performance* from the robot self-assessed lower *performance* after system use ($p = 0.007$, $F(1, 37) = 8.12$, $\eta^2 = 0.188$); and participants that perceived higher *social reciprocity* from the robot self-assessed lower *performance* after system use ($p = 0.025$, $F(1, 37) = 5.49$, $\eta^2 = 0.135$).

B. Behavioral Data

On average, participants interacted with the buzzer prompts 5.8 times ($SD = 2.5$) over 329 minutes of system use ($SD = 68$ min) and the robotic prompts 4.8 times ($SD = 2.7$) over 325 minutes of system use ($SD = 80$ min).

Based on evaluation metrics in the related literature, we

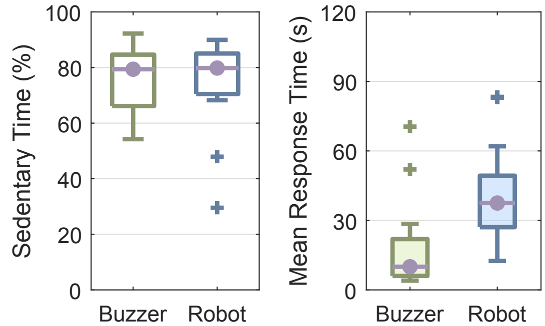

Fig. 4. *Left*: Sedentary times of each participant, expressed as a percentage of the total time enrolled in the study condition. *Right*: Mean response time after the start of the prompt for successful prompts (success defined as participant standing within 120 seconds of the start of the prompt).

identified sedentary time and standing prompt success as two key measures of break-taking system effectiveness. As shown in Fig. 4, there was no significant difference in sedentary time across conditions. There was also no significant difference between the two systems' abilities to encourage standing. Participants stood during prompts an average of 87.2% of the time (SD = 19.5%) when using the buzzer and 74.6% of the time (SD = 33.3%) when using the robot. An ANCOVA with system type as the independent variable and self-reported performance as the covariate showed that self-reported performance correlated with measured performance ($p = 0.002$, $F(1, 37) = 11.93$, $\eta^2 = 0.245$). Two participants did not stand during any of their robotic prompts.

Figure 4 also shows that participants responded significantly more quickly to the buzzer system compared to the robotic system ($p < 0.001$, $F(1, 33) = 21.83$, $\eta^2 = 0.364$). We noticed that in the recorded video of system prompts, participants would often watch or interact with the robot before standing, while participants typically responded to the buzzer immediately and without social interaction.

C. Video Analysis Data

OpenFace emotion and gaze variation results appear in Fig. 5. Participants' reduction in sadness was higher after interacting with the robot system compared to after interacting with the buzzer system ($p = 0.037$, $F(1, 35) = 5.13$, $\eta^2 = 0.129$). Change in happiness, anger, and gaze variation were not significantly different between conditions. One participant's video dataset was excluded due to the webcam failing to capture the participant's face and five videos were excluded under the criteria described in Section III-F.

While using this software, we noticed some inconsistencies and false positives in facial recognition. We checked the accuracy of OpenFace analysis by manually verifying OpenFace's identification of a human face in a randomly selected subset of 3.2% of analyzed frames, finding accuracy of $83.1 \pm 0.7\%$ with 95% confidence.

D. Interview Responses

The semi-structured interview yielded additional information on user experiences. When asked to select their preferred break-taking system, thirteen (13) participants selected the robotic system and six (6) selected the buzzer system.

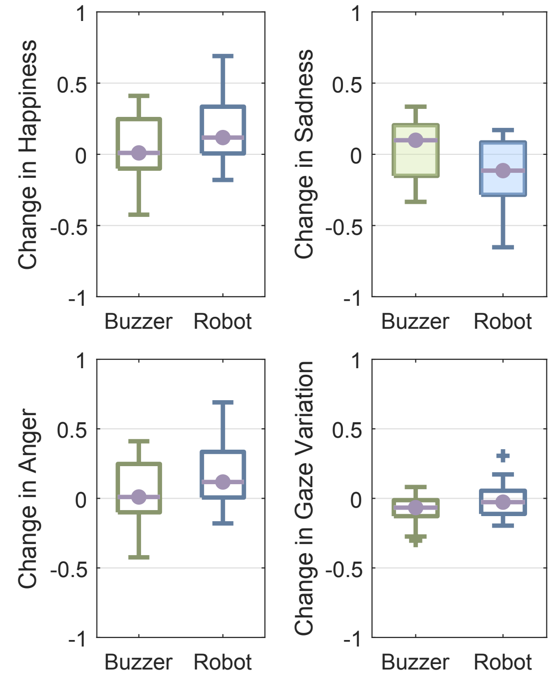

Fig. 5. Results of OpenFace analysis. Average emotion levels before and after the prompt had an output range from 0 to 5. Thus, axes for average emotion level change between before and after prompts range from -5 to 5. Axes for change in emotion level were truncated from -1 to 1 to illustrate differences. Gaze variation analysis showed similar distributions.

To further understand participant experiences, we used open coding on the interview data and identified 64 concepts for our codebook. Two reliability-trained coders labeled half of the interviews each, with a 10% overlap of labeled interviews to ensure the consistency of the codes. Using the equation described by Miles and Huberman [39], we computed an inter-rater reliability of 0.83, demonstrating a strong rater agreement. The codes that occurred most frequently appear in Table I. Key quotes and themes appear in our system design discussion in Section V.

E. Technical Challenges

To properly contextualize the overall results, it is important to note four key technical challenges encountered during the study. Firstly, system log errors led to incomplete chair sensor data for three study days. These log errors did not affect system functionality. For the behavioral analysis, we reconstructed missing data using video recordings for participant responses to break prompts and Fitbit data for sedentary time.

Two participants experienced a fatal buzzer system error. Electromagnetic interference from the vibrational motors on I^2C communications between the accelerometer and microcontroller led to continuous, unresponsive prompting. Since this error had the potential to negatively affect participant interactions with the buzzer, we recalculated the study statistics without these participants' data and found

TABLE I
INTERVIEW RESPONSE CODES THAT OCCURRED MOST FREQUENTLY.
▲ INDICATES BUZZER-RELATED CODES AND
● INDICATES ROBOT-RELATED CODES.

Description: Participant...	Instances
● Personified the robot (e.g. through pronouns, emotions)	39
Liked that the systems made them take breaks	22
● Felt that the interactions were confusing	21
● Described the robot as cute	18
▲ Thought that the buzzer interrupted their work	14
● Would use the robot for an extended time	13
● Thought the robot interrupted their work	13
▲ Liked that the buzzer was simple	12
● Felt that the robot was too distracting	11
● Thought the robot was not personalized enough	11
● Wanted the robot to move less	11
▲ Liked that the buzzer was easy to understand	11
▲ Felt that the buzzer was frustrating OR irritating	11
Mentioned headphones or earbuds while working	10
Wants the systems to have more sensing/data	9
● Described the robot as cool	9
▲ Would use the buzzer for an extended time	9
● Felt that the robot caught their attention well	8
● Felt that the robot took up too much desk space	8
● Wants the robot to charge itself	8

that omitting these participants did not change any results.

The Cozmo SDK setup limited guarantees of robotic system reliability. To control the robot, the system's Android phone needed to connect to the robot's ad-hoc network, and to connect, a user needed to press a button in the Cozmo application. Wireless instability and interference resulted in occasional disconnects, a fatal error as the robot could not be reconnected automatically. After each instance of this error, the robot was reconnected manually within 30 minutes.

Lastly, the Raspberry Pi was unable to provide enough current to charge the phone while the system was running. As a result, the phone had a battery life of approximately 6.5 hours in the study setup. This battery life limit resulted in three instances of the phone running out of battery, a fatal error for the robotic system. In one case, the participant did not notice the error, which occurred in the last hour of the session. In the other two cases, the problem was reported promptly and the research assistant brought a replacement system to continue the trial. We note that despite the occasional connection and battery life problems, participants preferred the robotic system overall.

V. DISCUSSION

Results supported the affective responses expected by **H1**. Compared to the buzzer system, the robotic system led to three significantly higher ratings of happiness, engagement, and enjoyment. In particular, effects on happiness and enjoyment had large effect sizes. Facial expression analysis supported survey results by showing that the robotic system led to a greater reduction of sadness. 68.4% of participants preferred the robotic system, and interview comments included many descriptions of the system as "fun," "cute," and "engaging." Interviews strongly featured personification of the robot, such as by assigning pronouns and attributing emotions and personality to the robot.

The break-prompting difference from **H2** was not supported, as user sedentary time and prompt success across systems was not significantly different. Although this result does not support the hypothesis, it suggests that the robotic system may reduce sedentary time as effectively as the buzzer system. Participants also tended to spend longer responding to robotic prompts. We explore some reasons below.

As a whole, the results show that participants felt positive about the robotic prompts. In interviews, participants frequently personified the robot, preferred the robotic system, and were willing to use the robotic system past the end of the study. However, this interest in robotic prompts occasionally worked counter to participant break-taking. One participant commented, "I felt like I was super motivated to sit in my seat [...] I was like, 'Oh, will this reset the timer? I want to see the robot move, so I'll sit a little longer.'" Another user noted that "it was kind of interesting to watch what it was doing. In the same way, I guess it kind of stopped me from getting up as immediately, because I wanted to see what it would do this time." These behaviors contributed to the higher response time for robot prompts.

At the same time, participants frequently overestimated the capability of the SAR system, leading to some disappointment. Participants desired more sensing and capabilities, such as one participant who said, "I wish that you could tell that right now I'm 100 percent focused and you would've waited." Another participant desired speech and facial recognition capabilities, commenting, "If it knows my name, if it can start speaking to me like, '[Participant], you need to take break [sic].' Or 'You've studied a lot.' [...] If it was looking to me [sic], then it would also be better." Results showed that higher *expected performance* and *social reciprocity* from the robot correlated negatively with self-reported break-taking performance, which emphasizes negative impacts from overestimating the robot. While participants responded positively about the robotic prompts, designing to minimize overestimation and ensuring expectations are met may improve results.

A. Design Implications

Our results indicate that future workplace SAR systems should strategically leverage positive user perceptions. In interview responses, participants gave praise for the expressiveness, cuteness, and physical embodiment of the robot. One participant noted, "I felt like it started to become kind of like a little pet on my desk, and I was happy." The interpretation of the robot as a pet demonstrates success in designing the robot as a social entity, and this metaphor can serve as a design tool for the future. In particular, emphasizing "cute" and "cool" aspects may be advantageous.

In addition, preconceptions of robot performance and sociability must align with the real capabilities of SAR systems. Several participants found the robot's behaviors to be "confusing" and "not personalized enough" and mentioned the desire for the robot to park and charge itself. The appearance of the SAR system led to expectations such as facial recognition, mobility, and object manipulation. Mismatched perceptions may lead to reduced performance perception of the SAR system. To understand likely preconceptions, we recommend surveying the target population about the robot and use case during the design phase and iterating the design based on feedback.

While participants broadly enjoyed the interactivity of the robot, conflicting user opinions on robot movement indicate the need for more personalization. Although ambient motion has improved perceptions of SAR systems in other roles [40], several participants disliked the built-in sleeping animation of the Cozmo robot, finding the motor noise and motion generated by this animation distracting. At the same time, others perceived the sleeping animation to be endearing. Overall, we recommend designing silent ambient behaviors with gradual, non-intrusive motion for users that may be distracted by more active ambient behaviors.

Participants also held differing opinions on the robot's prompt movements; one participant mentioned prompts "making it really engaging for me, getting me excited to actually get up and do something," while another thought the prompts were "a little too energetic [...] It was hard to know what [the robot] expected of me." Related worries about robot movement included worries about distracting co-workers or attracted unwanted attention. Future workplace SAR systems should allow for user feedback for system personalization, potentially including motion settings, appearance, and break intervals. For instance, the included smartphone system could request feedback after the first few interactions. Incorporating this element could move future work into the area of reinforcement learning.

The buzzer can also offer insights towards the design of SAR systems. Participants described the buzzer system as "simple," "minimalist," and "unobtrusive;" the buzzer system "does what it was supposed to." Some participants described the buzzer prompt as more sudden and as having an annoying, alarm-like quality that was likely to cause disuse over time, while other users actually preferred an annoying system as a means to force break-taking and noted that "[The robot] was maybe slightly not annoying enough." Our future SAR interventions require careful balance between prompt pleasantness and compellingness; an exceedingly pleasant prompt may fail from the start, while annoying prompt will lead to discontinued use of the system over time. Again, personalization and user feedback are necessary to achieve this balance.

Participants valued clear and immediate responses. One unique feature of the buzzer prompt was that, unlike the robot, it would stop vibrating when the user stood up and resume its prompting if a participant sat back down immediately. A user noted "So I liked that this one would go off, and then when you stand up and then sit immediately down, it would keep going off [...] it would force you to stay up for a long time." Implementing clearer robot responses to user behaviors (e.g., rewarding the user for standing) and better robot situational awareness (e.g., tracking user faces) can help to resolve the user uncertainty reflected in Table I and improve user perceptions of the SAR systems. In addition, this SAR system included designed behaviors that were only tested internally by the research group. As the focal point of the interaction, these behaviors should be tested and refined to a greater degree.

Lastly, several participants mentioned that the robot took up too much desk space, often due to the charging base of the robot. Minimizing the footprint of the system may also minimize both participant frustration and study facilitators' difficulties with deploying the SAR.

B. *Limitations and Future Work*

Limitations and design recommendations will drive our future work. We noted technical challenges such as connection stability and battery life in Section IV-E. To address these, we plan to bypass the need for a phone and connect the robot directly to our computer (e.g., by using an Android emulator), thereby establishing programmatic control of the wireless connection and removing phone battery issues in future workplace SAR system iterations.

Although our robotic system is already autonomous, adding the capabilities we recommend in Section V-A for feedback and personalization will help to promote system success in long-term use cases. We plan to develop additional prompt behaviors, design more meaningful robot feedback, create a system for user feedback, and implement real-time perception of user interruptibility to make the system more intelligent, independent, and adaptable. These improvements will reduce the gap between participant expectations and robot capabilities.

The limited number and background of participants in our work is not representative of all computer users. Our study also ran over a relatively short period of time, making the participants susceptible to system novelty. Because of the within-subject design of the study and the recruitment of students from our university, our results may include demand characteristics or please-the-experimenter bias.

However, this work provides design recommendations, indicates the most pertinent measures, and demonstrates the potential for SAR systems in this workplace, and thereby acts as a critical stepping-stone towards future work to move past these limitations and establish stronger results. With the justification and guidance provided here, we plan to follow up on this work with a longer-term, wider-sampling study with an improved SAR system.

C. *Conclusions*

This work presents a novel direct comparison between a SAR system and e-health-style intervention for reducing sedentary behavior in the workplace. We built upon previous SAR work to design a relatively complex workplace SAR system, and we found evidence that a SAR system is more satisfying than a non-social system. Results suggest that our work and general workplace SAR efforts should incorporate more robot adaptability, feedback, and independence.

These findings support the need for additional research and development for SAR systems in the workplace.

Acknowledgments

We thank Lilian Chan for her assistance developing Cozmo animations and Katelyn Swift-Spong for her advice.

References

[1] N. Owen, G. N. Healy, C. E. Matthews, and D. W. Dunstan, "Too much sitting: the population-health science of sedentary behavior," *Exercise and Sport Sciences Reviews*, vol. 38, no. 3, p. 105, 2010.

[2] D. Young *et al.*, "Sedentary behavior and cardiovascular morbidity and mortality: a science advisory from the american heart association," *Circulation*, vol. 134, no. 13, pp. e262–e279, 2016.

[3] C. Blehm, S. Vishnu, A. Khattak, S. Mitra, and R. W. Yee, "Computer vision syndrome: a review," *Survey of Ophthalmology*, vol. 50, no. 3, pp. 253–262, 2005.

[4] A. Hedge and S. J. Evans, "Ergonomic management software and work performance: an evaluative study," *Cornell University Human Factors Laboratory Technical Report/RP2501, Ithaca, NY*, 2001.

[5] D. W. Dunstan *et al.*, "Breaking up prolonged sitting reduces post-prandial glucose and insulin responses," *Diabetes Care*, vol. 35, no. 5, pp. 976–983, 2012.

[6] C. P. Mainsbridge, P. D. Cooley, S. P. Fraser, and S. J. Pedersen, "The effect of an e-health intervention designed to reduce prolonged occupational sitting on mean arterial pressure," *Journal of Occupational and Environmental Medicine*, vol. 56, no. 11, p. 1189, 2014.

[7] D. S. Bond, J. G. Thomas, H. A. Raynor, J. Moon, J. Sieling, J. Trautvetter, T. Leblond, and R. R. Wing, "B-mobile - a smartphone-based intervention to reduce sedentary time in overweight/obese individuals: A within-subjects experimental trial," *PLOS ONE*, vol. 9, no. 6, pp. 1–8, 06 2014.

[8] C. L. Edwardson *et al.*, "Effectiveness of the Stand More AT (SMArT) Work intervention: cluster randomised controlled trial," *BMJ*, vol. 363, 2018.

[9] D. W. Dunstan *et al.*, "Reducing office workers sitting time: rationale and study design for the Stand Up Victoria cluster randomized trial," *BMC Public Health*, vol. 13, no. 1057, 2013.

[10] W. A. Bainbridge, J. W. Hart, E. S. Kim, and B. Scassellati, "The benefits of interactions with physically present robots over video-displayed agents," *International Journal of Social Robotics*, vol. 3, no. 1, pp. 41–52, 2011.

[11] C. D. Kidd and C. Breazeal, "Robots at home: Understanding long-term human-robot interaction," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2008, pp. 3230–3235.

[12] C. G. Ryan, P. M. Dall, M. H. Granat, and P. M. Grant, "Sitting patterns at work: objective measurement of adherence to current recommendations," *Ergonomics*, vol. 54, no. 6, pp. 531–538, 2011.

[13] Y. Du, B. Liu, Y. Sun, L. G. Snetselaar, R. B. Wallace, and W. Bao, "Trends in Adherence to the Physical Activity Guidelines for Americans for Aerobic Activity and Time Spent on Sedentary Behavior Among US Adults, 2007 to 2016," *JAMA Network Open*, vol. 2, no. 7, pp. e197 597–e197 597, 07 2019.

[14] L. F. M. d. Rezende, M. Rodrigues Lopes, J. P. Rey-López, V. K. R. Matsudo, and O. d. C. Luiz, "Sedentary behavior and health outcomes: An overview of systematic reviews," *PLOS ONE*, vol. 9, no. 8, pp. 1–7, 08 2014.

[15] A. Must and D. Tybor, "Physical activity and sedentary behavior: a review of longitudinal studies of weight and adiposity in youth," *International Journal of Obesity*, vol. 29, no. S2, p. S84, 2005.

[16] D. P. Bailey, D. R. Broom, B. C. Chrismas, L. Taylor, E. Flynn, and J. Hough, "Breaking up prolonged sitting time with walking does not affect appetite or gut hormone concentrations but does induce an energy deficit and suppresses postprandial glycaemia in sedentary adults," *Applied Physiology, Nutrition, and Metabolism*, vol. 41, no. 3, pp. 324–331, 2016.

[17] T. Yates *et al.*, "Metabolic Effects of Breaking Prolonged Sitting With Standing or Light Walking in Older South Asians and White Europeans: A Randomized Acute Study," *The Journals of Gerontology: Series A*, 11 2018.

[18] T. M. Manini, L. J. Carr, A. C. King, S. Marshall, T. N. Robinson, and W. J. Rejeski, "Interventions to Reduce Sedentary Behavior," *Medicine & Science in Sports & Exercise*, vol. 47, no. 6, pp. 1306–13 010, 6 2015.

[19] A. Martin *et al.*, "Interventions with potential to reduce sedentary time in adults: systematic review and meta-analysis," *British Journal of Sports Medicine*, vol. 49, no. 16, pp. 1056–1063, 2015.

[20] M. J. Duncan *et al.*, "Balanced: a randomised trial examining the efficacy of two self-monitoring methods for an app-based multi-behaviour intervention to improve physical activity, sitting and sleep in adults," *BMC Public Health*, vol. 16, no. 1, p. 670, 2016.

[21] M. Fahim, T. Baker, A. Khattak, B. Shah, S. Aleem, and F. Chow, "Context mining of sedentary behaviour for promoting self-awareness using a smartphone," *Sensors*, vol. 18, no. 3, p. 874, 2018.

[22] H. op den Akker, V. M. Jones, and H. J. Hermens, "Tailoring real-time physical activity coaching systems: a literature survey and model," *User Modeling and User-Adapted Interaction*, vol. 24, no. 5, pp. 351–392, 2014.

[23] M. Monsey, I. Ioffe, A. Beatini, B. Lukey, A. Santiago, and A. B. James, "Increasing compliance with stretch breaks in computer users through reminder software," *Work*, vol. 21, no. 2, pp. 107–111, 2003.

[24] J. Kim, C. Cho, and U. Lee, "Technology supported behavior restriction for mitigating self-interruptions in multi-device environments," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, p. 64, 2017.

[25] D. Cooley and S. Pedersen, "A pilot study of increasing nonpurposeful movement breaks at work as a means of reducing prolonged sitting," *Journal of Environmental and Public Health*, vol. 2013, 2013.

[26] D. Feil-Seifer and M. J. Mataric, "Defining socially assistive robotics," in *9th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2005, pp. 465–468.

[27] J. Fasola and M. J. Matarić, "A socially assistive robot exercise coach for the elderly," *Journal of Human-Robot Interaction*, vol. 2, no. 2, pp. 3–32, 2013.

[28] N. T. Fitter, M. Mohan, K. J. Kuchenbecker, and M. J. Johnson, "Exercising with Baxter: Design and Evaluation of Assistive Social-Physical Human-Robot Interaction," in *Workshop on Personal Robots for Exercising and Coaching (PREC 2018)*, 2018.

[29] B. Scassellati *et al.*, "Improving social skills in children with asd using a long-term, in-home social robot," *Science Robotics*, vol. 3, no. 21, p. eaat7544, 2018.

[30] A. Tapus, M. Maja, and B. Scassellatti, "The Grand Challenges in Socially Assistive Robotics," *IEEE Robotics and Automation Magazine*, vol. 14, no. 1, p. N/A, 07 2007.

[31] S. Šabanović, S. M. Reeder, and B. Kechavarzi, "Designing robots in the wild: In situ prototype evaluation for a break management robot," *Journal of Human-Robot Interaction*, vol. 3, no. 1, pp. 70–88, 2014.

[32] A. Weiss, R. Bernhaupt, M. Tscheligi, D. Wollherr, K. Kuhnlenz, and M. Buss, "A methodological variation for acceptance evaluation of human-robot interaction in public places," in *Proc. IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2008, pp. 713–718.

[33] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[34] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): results of empirical and theoretical research," *Advances in Psychology*, vol. 52, pp. 139–183, 1988.

[35] T. Baltrusaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic action unit detection," *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 2015.

[36] P. Ekman, W. Friesen, and J. Hager, *Facial Action Coding System, Manual and Investigators Guide*, UT: Research Nexus, Salt Lake City, 2002.

[37] T. Baltrusaitis, "How to get the confidence value? Issue #270 TadasBaltrusaitis/OpenFace," Oct. 2017. [Online]. Available: https://github.com/TadasBaltrusaitis/OpenFace/issues/270

[38] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, N.J: L. Erlbaum Associates, 1988.

[39] M. B. Miles, A. M. Huberman, M. A. Huberman, and M. Huberman, *Qualitative data analysis: An expanded sourcebook*. sage, 1994.

[40] T. Kanda, M. Kamasima, M. Imai, T. Ono, S. Daisuke, H. Ishiguro, and Y. Anzai, "A humanoid robot that pretends to listen to route guidance from a human," *Autonomous Robots*, vol. 22, no. 1, p. 87, Sep 2006.