Designing and Validating Expressive Cozmo Behaviors for Accurately Conveying Emotions

Lilian Chan[1†], Brian J. Zhang[1†], Naomi T. Fitter[1]

***Abstract*— Robots have unique abilities to influence people, but when deploying robotic systems in assistive applications, roboticists must understand how users perceive these systems' behaviors. As part of an ongoing project to use robots as motivational break-taking aids, we present Cozmo behaviors that could function as the action space of a future robot learning strategy. Before deploying these behaviors in the wild, we evaluated them using an online video-based study with $N = 113$ participants. Results show that participant perceptions of Cozmo behaviors tend to match the intended valence and energy level. Furthermore, behavior valence in particular has a strong bearing on other perceived characteristics such as interaction appeal, trustworthiness, and safety. Facial expression and loudness acted as important covariates, which may help generalize these results to other behaviors and robots. The products of this work can benefit those who are interested in robot emotional expression and assistive robot applications.**

I. Introduction

Social robots possess a unique ability to engage and influence people using their embodiment in the physical world and deliberately chosen emotional behaviors [1]. Accordingly, such systems have gained adoption in spaces from entertainment to healthcare [2]. In our lab in particular, we are investigating the use of a Cozmo robot for encouraging workplace break-taking behaviors [3]. The Cozmo robot, which is an accessible and low-cost platform, has also been used to personalize tutoring [4], foster collaboration and inclusion [5], and facilitate in-home human interactions [6]. But in our project and other related domains, how can we understand human perceptions of the robot and its emotions? This work explores user evaluations of Cozmo's potential prompt behaviors as a preliminary step before incorporating such behaviors into a Markov Decision Process (MDP) model, in which Cozmo's action space will span typical emotional displays from human psychology research (e.g., [7]). Without adequate social skills, robots run the risk of ostracizing themselves from the people they are meant to assist [8]. Thus, research like the present investigation of robotic emotional display is essential to the adoption of social and socially assistive robots.

Previous efforts provide us with tools to conduct the proposed research on robot emotion and important evidence that the proposed application of Cozmo emotional behaviors is valid. Past studies of robots such as the EMYS [9], Roboceptionist [10], Geminoid F [11], KOBIAN [12], and Baxter [13] robots demonstrated the possibility of designing and validating specific emotional expressions for robotic systems. Many of these works draw design inspiration from the emotional models of Ekman [14] and Russell [7]. Past work also offers support of the proposed strategy to modulate Cozmo's displayed emotions to influence user behaviors; two related works show that people react differently depending on the mood of a Cozmo [6] or Roboceptionist [15] robot. Our work builds on these past efforts by studying a *richer array of Cozmo emotional behaviors* than considered by the most closely related work and by framing the understanding of Cozmo's emotions in a *new workplace context* that can support the success of this robot in a socially assistive role.

Fig. 1. A subset of Cozmo emotive behaviors investigated in this work.

To support the future successful application of emotional Cozmo behaviors as part of the robot's action space within an MDP model, we focus on the following research aims in the present paper: (1) understanding how well the intended and perceived valence and energy level of Cozmo behaviors align and (2) evaluating how differing Cozmo emotions affect other perceived characteristics of the robot (e.g., intelligence, safety, and trustworthiness). Towards these goals, we survey the most closely related literature in Section II. Section III outlines the methods of our online video-based study of Cozmo behaviors like those shown in Fig. 1. The results in Section IV show good alignment between the intended and perceived Cozmo emotions, and the findings also demonstrate that the emotional display of Cozmo influences other ways in which the robot is perceived by users. Section V discusses the key findings and design implications of this

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

[†]These authors contributed equally to this work.

work. Generally, the study results and included *open-source Cozmo behaviors* can benefit researchers who require well documented emotive Cozmo actions. Results of this study may also be applied to robots with similar features, such as those with LED display facial expressions. Further, the design process and methods used in our work can help to guide emotional categorization of new Cozmo behaviors and successful long-term applications of socially assistive robots like our break-taking Cozmo system.

II. Related Work

This research builds on past studies of human perception of robot emotion and social robots for long-term use.

A. Display and Perception of Robotic Emotions

Much of the past research on robot emotion builds from psychology works such as Ekman's basic emotions [14] and Russell's circumplex model of affect [7]. These references provide frameworks for how to categorize and characterize human emotions, thereby providing examples for how robot affective displays should be designed and modulated to mirror human feelings. Accordingly, several examples of past robotics research design and evaluate robot emotions based on Ekman's model (e.g., [9], [11], [16], [17]) and Russell's model (e.g., [10], [13], [18], [19]). Based on this past work and the potential of Russell's circumplex model to allow for a wider variety of emotional expressions, we based our designed Cozmo behaviors to fall at different levels on this model's valence and energy level axes.

Past studies also demonstrate promise for people's ability to accurately recognize what emotions a robot is attempting to convey. In the above-mentioned work, researchers typically design emotional expressions for robots and then evaluate how accurately onlookers can perceive the intended affect. Related work with the Roboceptionist platform suggested that people could correctly recognize different emotions and emotional intensities on this robot [10]. In a study of five basic emotions displayed by the Geminoid F robot, participants successfully identified happy, neutral, and sad facial expressions, but angry and fearful facial expressions had mixed results [11]. Fear was likewise the hardest emotion for participants to recognize in past studies of the EMYS [9] and EDDIE [18] robots. A study of custom facial expressions for the Baxter robot's face screen showed significant main effects of perceived robot pleasantness and energeticness, as well as viewer feelings of safety and pleasedness across the seven studied facial expressions [13]. Even with more minimalistic robots, emotion recognition is possible; a Roomba vacuum cleaning robot [19] and custom Maru robot [20] effectively displayed interpretable emotions using abstract cues like LED lights, motion/vibration, and sound.

The Cozmo robot studied in our work offers flexibility in emotional expression via its OLED screen, locomotion, head motion, and lift movements. Thus, we anticipate that viewers will be able to accurately interpret our designed emotional expressions for the robot. Compared to the built-in behaviors of the commercial Cozmo robot, we investigate more nuanced expression in this work; whereas the commercial robot mainly displays "happy," "sad," and "angry," we design and evaluate behaviors across the full span of Russell's circumplex model.

B. Social Robots for Long-Term Human-Robot Interaction

Beyond helping robots to display particular types of affect, emotional expression abilities can influence how people respond to social robots. For example, happy behaviors by a Cozmo robot typically move interactions with users forward, while sad behaviors interrupt interactions by causing users to look for problems in their earlier decisions [6]. Accordingly, the work presented in this paper has implications on social robots for long-term human-robot interaction (HRI).

Considering past long-term applications of social robots can help us design useful Cozmo behaviors for long-term interventions, such as our proposed break-taking support. Past long-term applications of these robots include HRI in healthcare, education, and domestic settings [21]. Past healthcare interventions such as older adult interactions with a PARO robot showed improvement in patient feelings and reductions of stress for patients and nurses [22]. Similarly, the Autom dieting support robot helped users to track significantly more calorie and exercise information than study participants who used computer- or paper-based logging systems [23]. In the education space, children who persisted in interactions with a Robovie robot improved their English skills more than peers who did not interact with the robot [24]. Deployments of the SPRITE and Jibo robots in home-based autism therapy interventions demonstrated potential for improving child numeracy [25] and social behavior [26] skills.

Across all of these long-term social robot applications, designed and perceived robot emotional expression can vastly change the interaction experience (as suggested by the results of [10] and [6]). Accordingly, the work presented in this paper contributes to long-term HRI efforts by providing clear and well understood emotionally expressive behaviors for the Cozmo robot. Precise emotional robot expressions in our envisioned break-taking application can help users to better understand the status of the robot and what they should do to perform well in the intervention.

III. Methods

The premise and motivation for this study arose from a larger effort to create a socially assistive robotic system for supporting break-taking at work. In pilot studies of this proposed intervention, participants sometimes had trouble understanding what the Cozmo robot was expressing [3]. Thus, the next steps of this project include this paper's efforts to more accurately design and display Cozmo emotions. The described mixed-design study was approved by the Oregon State University (OSU) Institutional Review Board under protocol #IRB-2019-0172.

A. Hypotheses

We developed three main hypotheses for the study:

H1: Human ratings of Cozmo's emotional behaviors will align with intended behavior valence and energy level on Russell's circumplex model [7]. Past robotics research supports the idea that perceptions of well designed robot emotions can fit this model [10], [13].

H2: Positively-valenced behaviors will lead to greater participant willingness to interact, opinion of robot trustworthiness, and interpretation of robot safety compared to negatively-valenced behaviors. This hypothesis was informed by pilot results from a small sample of test participants (with the same study design as in the present paper but a different initial test group of participants, before official data collection began) and is supported by past work like [15], in which people were more inclined to interact with a positive robot and rate their own affect as positive during the interaction than with a negatively-valenced robot.

H3: High-energy behaviors will lead to greater participant evaluation of robot intelligence compared to low-energy behaviors. This exploratory hypothesis was informed by the aforementioned pilot results.

B. Cozmo Behavior Design

Our Cozmo behaviors were created with Russell's circumplex model in mind [7]. We designed four animations for each of the following categories:

- Neutral Valence, High Energy Level (Arousal) [Active]
- Low Valence, High Energy Level [Unpleasant Active]
- Low Valence, Neutral Energy Level [Unpleasant] (e.g., Fig. 2)
- Low Valence, Low Energy Level [Unpleasant Inactive]
- Neutral Valence, Low Energy Level [Inactive]
- High Valence, Low Energy Level [Pleasant Inactive]
- High Valence, Neutral Energy Level [Pleasant]
- High Valence, High Energy Level [Pleasant Active] (e.g., Fig. 3)

This spread of emotional expression allows for the robot to imitate typical types of human affect and encourage users in different ways. For example, in the break-taking robot assistant scenario, a cooperative user may be sufficiently motivated by low-energy and neutral- or positive-valenced cues, while a stubborn user may not be spurred to take a break until Cozmo displays a negatively-valenced and high-energy behavior.

Eight behavioral categories with four behaviors in each category were chosen to provide adequate coverage of Russell's circumplex model [7]. We used the Cozmo software development kit (SDK) [27] to hand-design the 32 animations from a combination of built-in behaviors and custom actions. The video included with this paper shows a demonstration of one behavior from each category. Open-source versions of the resulting behaviors are available in the code repository associated with this paper [28]. The animations leveraged Cozmo's facial expressions, locomotion, head motion, and lift movement. Reducing noise in workplace robotic interventions is essential [29], so we muted Cozmo's audio (e.g., built-in non-linguistic utterances and speech) for all of the

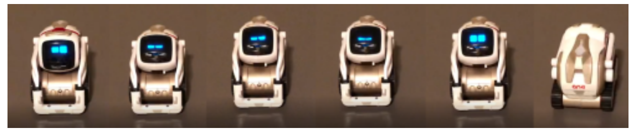

Fig. 2. A subset of cropped frames from the "frustrated" behavior.

Fig. 3. A subset of cropped frames from the "celebratory" behavior.

animations via the Cozmo phone application settings. As observable in the video included with this paper, consequential sounds, such as the noise of Cozmo's motors, were still audible in the animation videos. Each animation was between 12 and 18 seconds long (M = 13.84, SD = 1.65). Informal piloting with lab members, peers, and animation expert Carmen Tiffany helped to ensure that the programmed behaviors were a reasonable fit with the intended affective categories. Videos of the resulting behaviors appear in the code repository associated with this work [28].

C. Measurement

The survey collected information and responses from participants with the following questionnaires:

Attitudes questionnaire: at the beginning of the survey, the Negative Attitudes towards Robots Scale (NARS) captured participants' preconceptions on robots for potential use as a covariate in subsequent analyses. Participant ratings of agreement with fourteen statements on seven-point Likert scales were combined into subscales of negative attitudes towards *interactions with robots*, *social influence of robots*, and *emotions in robots* as described in [30].

Demographic questionnaire: next, participants answered demographic and occupational questions.

Post-stimulus questionnaire: after each stimulus, the survey asked how much participants agreed with the following statements on seven-point Likert scales:

1) This robot seems pleasant.
2) This robot seems energetic.
3) I would interact with a robot that behaves this way.
4) A robot that behaves this way seems trustworthy.
5) A robot that behaves this way seems safe to interact with.
6) A robot that behaves this way seems intelligent.

Questions 1 and 2 measure *valence* and *energy level* of Russell's circumplex model [7], while questions 3-6 helped to enrich the data without overburdening participants through measurements selected from common metrics for social robots and the Godspeed survey [31], [32]. Lastly, participants indicated which of the following robot feature(s) most influenced their responses to the preceding questions: "facial expressions," "locomotion of robot," "other robot motion (head, lift)," "sounds of robot," or "other."

Overall perceptions questionnaire: after responding to one stimulus from each category (eight total stimuli), participants answered questions about their envisioned relationship(s) with Cozmo, the robot's perceived gender, and metaphor(s) for the robot's behavior. These items arose from our discussions with Carmen Tiffany, during which we realized that understanding how users naturally perceive the robot will be important to successful design of robot break-taking interventions.

Free-response question: at the close of the study, an optional free-response question asked participants to describe the characteristics of the stimuli that most strongly influenced the survey responses.

We also extracted objective measures to help generalize the results of this work:

Facial expression measures: for each stimulus, a research team member coded the proportions of video frames in which the robot displayed a positive, neutral, or negative facial expression. Mappings from facial expression to valence were accomplished using facial expression-to-emotion keys released by Anki in their original robot promotional materials and the associated placement of these emotions on Russell's circumplex model. In the future, this data could be extracted programmatically by tracking which faces are used and for how long in any given animation.

Audio measures: N5 loudness and peak loudness were extracted from the audio of each stimulus [33]. We also calculated an average fundamental frequency, derived by weighting fundamental frequencies found using normalized correlation functions [34] with corresponding loudness for subsections of the video. As movement of the Cozmo robot results in servo noises, we suspected that these values might relate to energeticness of robot behaviors.

D. Participants

The study was completed by $N = 113$ students from Oregon State University between 18 and 45 years of age ($M = 22.0$, $SD = 5.9$), including 79 cisgender women and 34 cisgender men. While most participants (66.4%) were pursuing a major course of study in science, technology, engineering, or mathematics (STEM), participants reported little to no experience with both robots ($M = 1.75$, $SD = 0.74$) and Cozmo ($M = 1.09$, $SD = 0.39$) on a seven-point Likert scale. Participant responses on the NARS assessment [30] showed generally neutral attitudes towards *interactions with robots* ($M = 3.53$, $SD = 1.09$), *social influence of robots* ($M = 4.58$, $SD = 1.14$), and *emotions in robots* ($M = 4.53$, $SD = 1.21$).

E. Procedure

Participants were recruited through a university student pool, through which introductory psychology course students can complete studies for class credit. Consenting participants completed the study online using a Qualtrics survey. The initial survey pages included questions about the participant's general views of robots and demographic information, followed by an introduction to the study. Participants then viewed 8 randomly assigned Cozmo animations and answered post-stimulus questions about the robot's valence, energy level, other characteristics of interest, and key aspects of the robot's behavior. Stimulus assignment was balanced across respondents. Lastly, participants answered closing questions about general perceptions of Cozmo. To ensure data quality, we required responses to all questions except open-ended free response fields. Timers throughout the survey helped to ensure that participants watched the full robot videos and spent sufficient time completing question sets, and attention-check questions helped to exclude any respondents who randomly selected answers.

F. Analysis

As participants responded to a subset of the stimuli, survey responses were analyzed through linear mixed models (LMMs) with $\alpha = 0.05$ and participant number as the random effect to account for participant-specific response patterns. Where applicable, significant differences between behavior categories were further explored using the Holm-Bonferroni method, which identified pairwise differences and the resulting p values of each comparison [35]. We report effect size via marginal r^2, the variance explained by the fixed effects of the LMM, where $r^2 = 0.010$ is considered a small effect, $r^2 = 0.040$ a medium effect, $r^2 = 0.090$ a large effect, and $r^2 = 0.016$ a very large effect [36].

Based on the objective measures, we identified useful objective scales via an exploratory feature engineering process. We then applied those objective scales as fixed effects in LMMs as described above.

Results were analyzed using jamovi [37], [38], [39].

IV. Results

Key results include the locations of Cozmo behaviors on pleasantness and energy level axes, differences in robot characteristic ratings between behavior categories and along objective metrics, and general perceptions of Cozmo that can help inform how we use this robot in future interventions.

A. Response Alignment with Russell's Circumplex Model

LMMs using behavior category as the fixed effect indicated that the categories led to differences in perceived *valence* ($p < 0.001$, $F(7, 1707.07) = 76.17$, $r^2 = 0.190$) and *energy level* ($p < 0.001$, $F(7, 1712.70) = 165.78$, $r^2 = 0.353$). Post-hoc analysis for *valence* showed that 23 of 28 pairwise comparisons yielded significant differences, with 18 being very significant ($p < 0.001$). Differences between Active and Pleasant Active, Active and Pleasant Inactive, Pleasant Active and Pleasant Inactive, Unpleasant and Unpleasant Inactive, and Pleasant and Pleasant Inactive were not significant. Similarly, post-hoc analysis for *energy level* showed that 25 of 28 pairwise comparisons had pairwise differences, with 24 being very significant ($p < 0.001$). Differences between Active and Unpleasant Active, Inactive and Unpleasant Inactive, and Unpleasant Active and Pleasant were not significant. The placement of each animation along these two axes appears in Fig. 4.

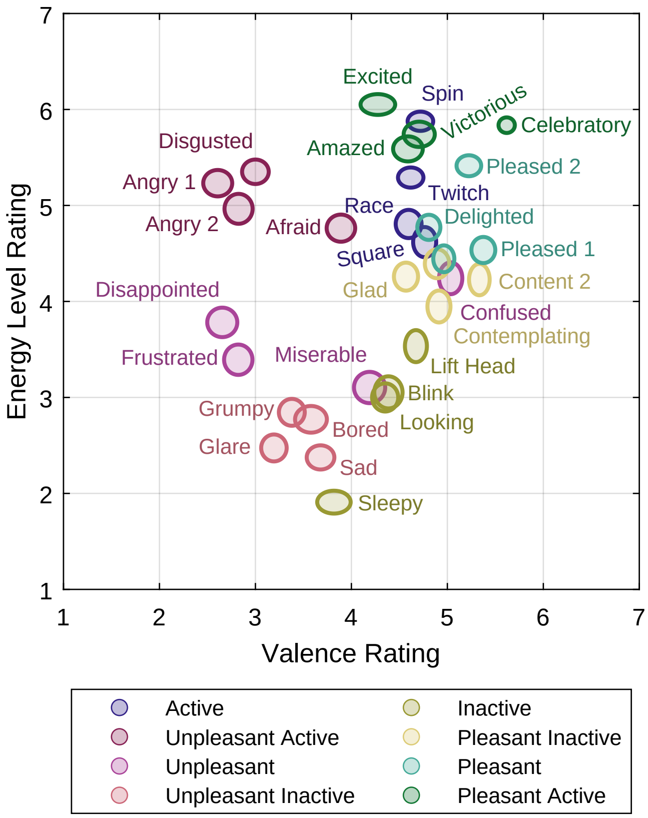

Fig. 4. Placement of the 32 emotional Cozmo behaviors on the valence and energy level axes from Russell's circumplex model. Centroids of the ovals fall at the coordinates of mean rating on each scale, and the width and height of each oval are proportional to the one-tenth of the standard deviation of that behavior's valence and energy level ratings, respectively. The "Happy" (Pleasant) and "Content 1" (Pleasant Inactive) ovals remain unlabeled for visibility.

B. *Influence of Behavior Categories on Robot Perception*

The LMMs also showed that behavior categories correlated with significant differences in perceived *willingness to interact* ($p < 0.001$, $F(7, 1697.97) = 43.26$, $r^2 = 0.096$), *trustworthiness* ($p < 0.001$, $F(7, 1707.07) = 76.17$, $r^2 = 0.104$), *safety* ($p < 0.001$, $F(7, 1707.07) = 76.17$, $r^2 = 0.100$), and *intelligence* ($p < 0.001$, $F(7, 1707.07) = 76.17$, $r^2 = 0.032$). Figure 5 shows the response distributions for each category and the results of pairwise comparisons.

At a high level, low-valence behaviors led to a significant reduction in willingness to interact, trustworthiness, and safety compared to all other behaviors. In all four measures, participants rated Pleasant behaviors highest and, in all measures except intelligence, participants rated Unpleasant Active behaviors lowest. Some significant differences appeared between category ratings for robot intelligence but did not form a clear overall trend.

C. *General Perceptions of the Cozmo Robot*

At the end of the post-stimulus questionnaire, participants checked the most important factor(s) that influenced their responses to the six previously mentioned post-stimulus questions. Facial expressions were checked 35.7% of the time, overall locomotion 29.1% of the time, head and lift motion 24.7% of the time, sound 9.6% of the time, and "other" 1.0% of the time.

The close of the survey also asked participants to imagine their relationship with the robot if they had the robot in their home for day-to-day use. When checking their relationship(s), "toy" occurred 48.0% of the time, "pet" 26.0% of the time, "friend" 14.5% of the time, "peer" 4.6% of the time, "child" 3.5% of the time, and "other" 3.5% of the time. Participants also described the robot as "male" (59.3%), "androgynous" (15.9%), "no gender" (8.9%), "male-androgynous" (8.0%), or "unsure" (8.0%). Lastly, when asked what the robot behaved like, participants checked "mammal" 38.0% of the time, "[not] like any animal" 35.7% of the time, "bird" 3.9% of the time, "invertebrate" 3.1% of the time, "reptile" 2.3% of the time, "amphibian" 2.3% of the time, "fish" 0.8% of the time, and "other" 14.0% of the time.

D. *Exploratory Feature Engineering of Objective Measures*

A correlation matrix of the six objective measures indicated three potential groupings of the measures: (1) FE, a facial expression scale, formed from the proportions of positive, neutral, and negative facial expressions, with negative facial expressions correlating negatively with the other expression types, (2) LS, a loudness scale formed from peak loudness and N5 loudness, and (3) FR, a frequency scale formed from the average fundamental frequency. In particular, neutral expression proportion and negative expression proportion correlated strongly and negatively ($r = -0.815$, $p < 0.001$), while peak loudness and N5 loudness correlated strongly and positively ($r = 0.825$, $p < 0.001$).

As the expression proportions generally combined to reach 1 and neutral and negative expressions correlated so strongly, we simplified the facial expression scale into a two-element scale like so: $FE = P - N$, where P is the proportion of positive expressions and N is the proportion of negative expressions. Similarly, the loudness scale was formed as: $LS = PL + NL$, where PL represents peak loudness and NL represents N5 loudness. For LS and FR, the frequency scale, we standardized the scales so that $M = 0$ and $SD = 1$ to be able to compare parameter estimates.

E. *Exploring Objective Scales as Model Effects*

LMMs with all three objective scales as the fixed effects showed that the frequency scale yielded F values and parameter estimates between 1 and 3 orders of magnitude smaller than the other scales and null results for 3 of 6 LMMs. Thus, we excluded the frequency scale from the final analysis.

LMMs with the facial expression and loudness scales as fixed effects showed that these two scales acted as important covariates for all measures except intelligence. Figure 6 shows the position of each behavior along these scales.

The results of the LMMs yielded significance for both fixed effects with $p < 0.001$ except for the effect of the loudness scale on perceived *intelligence*, which was not significant. Overall LMM effect sizes were more than very large for *valence* ($r^2 = 0.180$) and *energy level* ($r^2 = 0.304$),

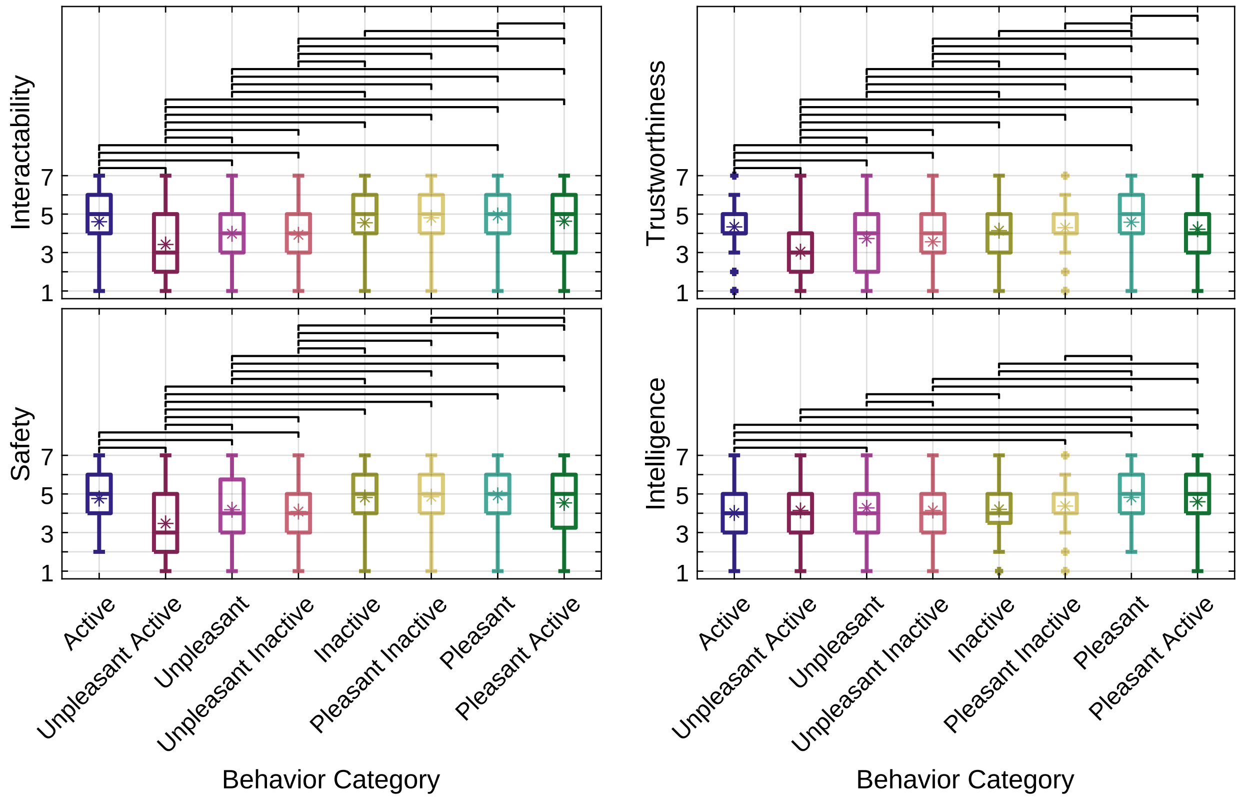

Fig. 5. Post-stimulus RoSAS responses for Studies 3-4. Boxplots include boxes from the 25th to the 75th percentiles, center lines for medians, asterisks for means, whiskers up to 1.5 times the inter-quartile range, and "+" marks for outliers. Brackets above the boxplots indicate significant pairwise differences.

approximately large for *willingness to interact* ($r^2 = 0.087$), *trustworthiness* ($r^2 = 0.089$), and *safety* ($r^2 = 0.100$), and less than small for *intelligence* ($r^2 = 0.007$).

For all measures except energy level, parameter estimates showed a positive and more significant effect for the facial expression scale and a negative and smaller effect for the loudness scale. For *energy level*, the parameter estimates for both scales were positive, but loudness had a larger and more significant parameter estimate, indicating a stronger positive relationship between loudness and energy level.

V. Discussion

H1 was well supported by the data. We saw a general trend for our designed affective Cozmo behaviors to fall as expected on Russell's circumplex model. Two of the behavior categories were most challenging to design as intended: Pleasant Inactive and Active. This is consistent with past work in [13], which also had difficulty designing "calm"- and "alert"-seeming robot behaviors. In both our work and this past paper, particularly for behaviors with high valence and/or high energy levels, ratings of the two characteristics tended to be directly related. Participant free response feedback supported the idea that Cozmo demonstrated a representative span of emotions; participants commented that Cozmo was "able to portray a variety of moods" and that Cozmo's behavior "resembled emotions such as energetic, sad, mad, or happy."

The results supported **H2** since behaviors that were designed to be higher- and neutral-valenced typically led to higher ratings of willingness to interact with the robot, robot trustworthiness, and robot safety compared to lower-valenced behaviors. The majority of significant pairwise differences appeared between low-valence behaviors and other valence level categories. Thus, it appears that the distinction between "negative" and any other valence level is more important than other differences in the three total levels considered, perhaps because only negative behavior seems alienating or intimidating. Some participant comments, such as "I would feel more comfortable interacting with a robot that had [...] happy facial characteristics," "when the robot had [a] happier facial expression I considered it more trustworthy and safe," and "angry facial expressions made me feel that the robot may not be trustworthy" support this interpretation. Furthermore, strong correlations between the exploratory facial expression scale and all measures indicate that the comments focus on facial expressions may reflect a consistent underlying effect.

H3 was not well supported by our results. Some differences appeared in the intelligence ratings, but they did not seem strongly related to energy level. Six of the significant pairwise differences were between high-energy categories and other behavior types, but the direction of the relationship was not uniform. We think this result may mean that a high-energy robot might seem playful (e.g., "egging you on to play with them") or angry/uncontrollable/aggressive (e.g., "it acted crazy, it was not very pleasant," "[it] seems to throw fits") depending on the circumstances. Low-energy behavior perceptions were also split; respondents saw these behaviors

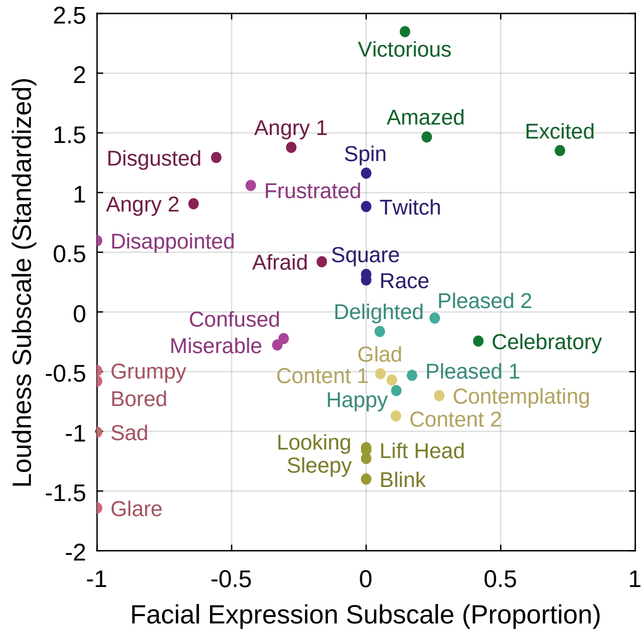

Fig. 6. Placement of the 32 emotional Cozmo behaviors on the loudness and facial expression subscales.

as either shifty (e.g., "it created a feeling of uncertainty/being watched," "it felt hostile or suspicious," "it seemed lazy it made me not want to trust it") or pleasant (e.g., "I prefer more subtle or smooth movements," "I thought I could trust the [slower] robot"). Exploratory analysis using the loudness scale, which may act as a proxy variable for robot motion speed and amplitude, did not show a significant effect on intelligence. However, with the exception of energy level, the loudness scale correlated with a decrease in the other five ratings. Thus, behaviors that moderate the amount of motion may successfully increase perceived interactability, trustworthiness, and safety. Further investigation is required to determine whether the motion or the sound is responsible for these effects.

Based on the closing survey questions, we learned more about how participants viewed Cozmo generally and how participants viewed their relationship with Cozmo. One of the most resounding agreements was that Cozmo presents as male; more than half of respondents identified Cozmo's gender as male, and no one placed the robot on the female side of the spectrum. Cozmo facial expressions seemed to be the single most important attribute for participant perceptions, although robot base, head, and lift motion also played a notable role. Popular interpretations of the robot's relationship with the user were as a toy or pet, and most responses indicated that the robot was either like a mammal or like no animal at all. This split seems to indicate that when adapting the robotic system for break-taking support in the workplace, we might be wise to design two main types of interaction paradigms: one which is more pet-like (with the design metaphor of how a dog or cat might behave), and one that is more electronic device-like (with design metaphors of gadgets, desk toys, or phone applications).

A. *Design Implications*

Our results indicate that we can produce different emotions that are interpreted by people roughly as intended around the entirety of Russell's circumplex model. This result helps to push the boundary on the primarily happy or sad emotions built into the commercial Cozmo robot. Notably, the provided open-source behaviors accomplish these differences in perception even without the non-linguistic utterances and other added sounds included in many of Cozmo's built-in behaviors. We also found that we can change other perceived characteristics of the robot (e.g., participant desire to interact with Cozmo, perceived trustworthiness, and perceived safety) just by changing the affective display of the robot. For break-taking, these discoveries will reduce potential confusion arising from break-taking prompts and increase participants' willingness to use the robotic break-taking system.

Additional insights from the study help us to identify design paradigms for future human-Cozmo interactions in this robot's various use cases (e.g., habit formation, entertainment, social play facilitation, tutoring), as well as the objective features of facial expression proportions and loudness that may be manipulated to achieve desired perceptions. These objective features may be extended to other robots' facial expressions and expressive motions to support behavior design and improve the match between intended and actual perceptions. Thus, the end design products of this work provide a rich emotional action space for use in the proposed future Cozmo MDP, offer ways to change how the user perceives and responds to Cozmo and similar robots through modulations of the presented robot emotions alone, and inform the design of robot interactions or relationships with people to support successful future interventions.

B. *Key Strengths and Limitations*

One strength of this work is the involvement of an animation expert in the design of the Cozmo behaviors. This collaboration may have supported the positive results, in which the Cozmo behaviors were perceived mostly as intended. This project supports concrete future applications of the Cozmo system and models good design for social robotic deployments (i.e., understanding how robot behaviors are perceived in isolation before introducing them in a more complex scenario). Additionally, the open-sourcing of the involved Cozmo behaviors [28] supports progress of social robotics as a whole by contributing to the collective resources for a common/accessible robot platform.

A main limitation of this work is its use of an online video-based study (rather than in-person interactions) to evaluate Cozmo behaviors. Our planned extension of this work will involve in-person studies to help confirm that the results generalize to in-the-wild interactions with robots. The participant population also did not fully represent all potential users of our robotic system. Based on our recruitment pool, most participants were young adult university students from the United States. In our follow-up work, we plan to recruit participants with a wider variety of backgrounds and identities to better represent potential Cozmo users.

VI. Conclusions

In this paper, we present the evaluation of Cozmo behaviors for use in a future assistive robotic system. We designed these behaviors with the help of a collaborator from animation, and we evaluated the Cozmo emotional expressions using an online video-based survey. Results of the survey showed that the behaviors generally matched their intended perception and provided important design information (e.g., a more positively-valenced robot is perceived better in other user ratings, and Cozmo is generally viewed as a male pet-like or toy-like entity). These findings can support the introduction of our designed Cozmo behaviors into future systems such as our proposed break-taking aid. Other researchers and robot designers with interest in social robotics, socially assistive robotic interventions, and open-source tools for affective robots can benefit from this work.

VII. Acknowledgments

We thank Carmen Tiffany for feedback on the robot animations and Christopher Sanchez for data collection support.

References

[1] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 143–166, 2003.

[2] C. Breazeal, A. Takanishi, and T. Kobayashi, "Social robots that interact with people," *Springer Handbook of Robotics*, pp. 1349–1369, 2008.

[3] B. J. Zhang, R. Quick, A. Helmi, and N. T. Fitter, "Socially assistive robots at work: Making break-taking interventions more pleasant, enjoyable, and engaging," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 11 292–11 299.

[4] D. Leyzberg, S. Spaulding, and B. Scassellati, "Personalizing robot tutors to individuals' learning differences," in *Proc. of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2014, pp. 423–430.

[5] S. Gillet, W. van den Bos, and I. Leite, "A social robot mediator to foster collaboration and inclusion among children," in *Proc. of Robotics: Science and Systems (RSS)*, 2020.

[6] H. R. Pelikan, M. Broth, and L. Keevallik, "'Are you sad, Cozmo?' How humans make sense of a home robot's emotion displays," in *Proc. of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2020, pp. 461–470.

[7] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[8] K. Dautenhahn, "Socially intelligent robots: Dimensions of human-robot interaction," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1480, pp. 679–704, 2007.

[9] J. Kędzierski, R. Muszyński, C. Zoll, A. Oleksy, and M. Frontkiewicz, "EMYS-emotive head of a social robot," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 237–249, 2013.

[10] R. Kirby, J. Forlizzi, and R. Simmons, "Affective social robots," *Robotics and Autonomous Systems*, vol. 58, no. 3, pp. 322–332, 2010.

[11] C. Becker-Asano and H. Ishiguro, "Evaluating facial displays of emotion for the android robot Geminoid F," in *Proc. of the IEEE Workshop on Affective Computational Intelligence*, 2011, pp. 1–8.

[12] M. Zecca et al., "Whole body emotion expressions for KOBIAN humanoid robot: Preliminary experiments with different emotional patterns," in *Proc. of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2009, pp. 381–386.

[13] N. T. Fitter and K. J. Kuchenbecker, "Designing and assessing expressive open-source faces for the Baxter robot," in *Proc. of the International Conference on Social Robotics (ICSR)*. Springer, 2016, pp. 340–350.

[14] P. Ekman, W. V. Friesen, and P. Ellsworth, *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. Pergamon, 1972.

[15] R. Gockley, J. Forlizzi, and R. Simmons, "Interactions with a moody robot," in *Proc. of the ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, 2006, pp. 186–193.

[16] P. Ponce, A. Molina, and D. Grammatikou, "Design based on fuzzy signal detection theory for a semi-autonomous assisting robot in children autism therapy," *Computers in Human Behavior*, vol. 55, pp. 28–42, 2016.

[17] J. Li and M. Chignell, "Communication of emotion in social robots through simple head and arm movements," *International Journal of Social Robotics*, vol. 3, no. 2, pp. 125–142, 2011.

[18] S. Sosnowski, A. Bittermann, K. Kuhnlenz, and M. Buss, "Design and evaluation of emotion-display EDDIE," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2006, pp. 3113–3118.

[19] S. Song and S. Yamada, "Designing expressive lights and in-situ motions for robots to express emotions," in *Proc. of the International Conference on Human-Agent Interaction*, 2018, pp. 222–228.

[20] ——, "Expressing emotions through color, sound, and vibration with an appearance-constrained social robot," in *Proc. of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2017, pp. 2–11.

[21] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: A survey," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.

[22] K. Wada, T. Shibata, T. Saito, and K. Tanie, "Effects of robot-assisted activity for elderly people and nurses at a day service center," *Proc. of the IEEE*, vol. 92, no. 11, pp. 1780–1788, 2004.

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

[24] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro, "A two-month field trial in an elementary school for long-term human–robot interaction," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 962–971, 2007.

[25] C. E. Clabaugh et al., "Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders," *Frontiers in Robotics and AI*, vol. 6, p. 110, 2019.

[26] 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, 2018.

[27] Anki, "Cozmo SDK (Version 1.4.10)[Computer software]," 2019.

[28] L. Chan and N. T. Fitter. (2020) expressive_cozmo_behaviors repository. [Online]. Available: https://github.com/shareresearchteam/expressive_cozmo_behaviors

[29] 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.

[30] D. S. Syrdal, K. Dautenhahn, K. L. Koay, and M. L. Walters, "The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study," in *Proc. of the AISB Symposium on New Frontiers in Human-Robot Interaction*, 2009.

[31] A. Steinfeld et al., "Common metrics for human-robot interaction," in *Proc. of the ACM SIGCHI/SIGART Conference on Human-Robot Interaction (HRI)*, 2006, p. 33.

[32] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 71–81, Jan. 2009.

[33] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*, 3rd ed., ser. Springer Series in Information Sciences. Berlin, NY: Springer, 2007, no. 22.

[34] B. S. Atal, "Automatic speaker recognition based on pitch contours," *The Journal of the Acoustical Society of America*, vol. 52, no. 6B, pp. 1687–1697, Dec. 1972, publisher: Acoustical Society of America.

[35] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.

[36] D. C. Funder and D. J. Ozer, "Evaluating effect size in psychological research: Sense and nonsense," *Advances in Methods and Practices in Psychological Science*, vol. 2, no. 2, pp. 156–168, Jun. 2019.

[37] The jamovi project, "jamovi (Version 1.6) [Computer software]," 2020.

[38] R Core Team, "R: A language and environment for statistical computing (Version 4.0) [Computer software]," 2020.

[39] M. Gallucci, "GAMLj: General analyses for linear models [jamovi module]," 2019.