Backseat Teleoperator: Affective Feedback with On-screen Agents to Influence Teleoperation

Daniel J. Rea, James E. Young
University of Manitoba
{daniel.rea, young}@cs.umanitoba.ca

***Abstract*—We investigate if an on-screen agent that reacts to a teleoperator's driving performance (e.g., by showing fear during poor driving) can influence teleoperation. Serving as a kind of virtual passenger, we explore if and how this agent's reactions may impact teleoperation. Our design concept is to create an emotional response in the operator (e.g., to feel bad for the agent), with the ultimate goal of shaping driving behavior (e.g., to slow down to calm the agent). We designed and implemented two proof-of-concept agent personas that react differently to operator driving. By conducting an initial proof-of-concept study comparing our agents to a base case, we were able to observe the impact of our agent personas on operator experience, perception of the robot, and driving behavior. While our results failed to find compelling evidence of changed teleoperator behavior, we did demonstrate that emotional on-screen agents can alter teleoperator emotion. Our initial results support the plausibility of passenger agents for impacting teleoperation, and highlight potential for more targeted, ongoing work in applying social techniques to teleoperation interfaces.**

Keywords — teleoperation; human-robot interaction; social interfaces

I. Introduction

Robot teleoperation is becoming part of more industries and available to more people for everyday use. Robots are used to inspect factories, save people in urban disasters, or perform remote medical procedures. Non-specialists now also use teleoperated robots more frequently to attend work from home, participate in conferences, and even visit museums and attractions in other cities. While teleoperation interfaces have improved, safely operating a robot is difficult, often due to the challenge of understanding the state of the robot and the surrounding environment [1]–[3]. Thus, while mistakes such as collisions have been shown to often be due to operator error [4], [5], researchers and companies have tried to improve operator performance by improving interfaces that support safe operation.

We present an approach to improving teleoperator performance that aims to use social feedback to impact an operator's mental state, with the ultimate goal of trying to shape how they drive the robot. Specifically, we add an interactive agent to a simple teleoperation interface, like a virtual passenger, which reacts to operator driving using emotional feedback. Ideally, the operator may feel empathy and compensate by altering their driving (Fig. 1). For example, if the agent acts scared following a collision, the operator may feel empathy and automatically drive more safely to console the agent. This effect, of a person witnessing an emotion and, in response, changing their behavior or feeling an emotion themselves, is well-established in other fields (e.g. [6]–[8]). In this paper, we present and explore a proof of concept using this technique to shape teleoperator experience with the intent to change operation behavior.

a)
b)

**Fig. 1a) An on-screen "virtual passenger" agent reacts to poor driving by exhibiting anxiety, with the intention of impacting the teleoperation experience
b) the interface displayed during robot teleoperation.**

It has become common to explore how robots can use human- or animal-like social communication techniques when working with people, in an attempt to improve and simplify communication with them [9]. For example, autonomous robots co-located with people can use techniques such as expressive movement [10], gaze [11], or even animal-like tail movements to convey robotic state or intention [12]. However, apart from social tele-operation (where a robot is a proxy for two remote people interacting), there has been little work done that explores how a tele-operated robot can similarly use social techniques to support their operators. As such, we present this paper as a proof of concept, where a tele-operated robot aims to use techniques from social HRI to impact the tele-operator.

We designed and implemented a virtual passenger on the tele-operated robot which reacts to the operator's driving (e.g., average speed, collisions) in real time by displaying an emotion (Fig. 1), with the goal of shaping the tele-operator experience and perhaps ultimately their driving behavior. For example, if the agent reacts with a positive emotion, such as smiling, the operator may similarly become more positive which may reinforce the current driving style. Conversely, a negative agent emotion, such as anxiety or fear, might discourage the current driving behavior. To explore this approach, we designed two agent variants, each using a different affect feedback model, and conducted an initial study to investigate how these agents may impact the teleoperation experience and operator's driving.

Our results indicate that affective feedback passenger agents can create emotional change in teleoperators. However, we found no compelling evidence that they changed driving behavior in this case; our analysis highlights limitations and avenues for improving both the agent and study design that will be useful for follow-up work. Overall, our work serves as a proof of concept of using affective feedback-based interfaces in teleoperation, and of using social interaction techniques to support operators in general, which we envision will be an important research topic for teleoperation moving forward.

II. Related Work

A goal of teleoperation researchers is to help people use robots to accomplish tasks, such as inspection or search-and-rescue tasks. One way to improve how well operators perform these tasks is to reduce critical incidents such as collisions, increase awareness, and reduce task completion time [3], [13], [14]. Techniques may do this through novel interfaces to better communicate information about the robot (e.g., [15], [16]) and its surroundings (e.g., [13], [17], [18]). Teleoperation mistakes may also be avoided by making it easier to control the robot (e.g., [18]–[21]), or leverage psychology to improve the operator's behavior [3], [22], [23]. In this work, we explore how social signals may also be employed by teleoperation interfaces to shape operator experience and performance during teleoperation.

Research in traffic psychology has shown that a driver's psychological state can change their driving behavior [24], [25]. These changes may be due to the perception of the vehicle itself [26], the surrounding environment [27], the driver's mood [28], or even the physical controls of the vehicle [29], [30]. This body of work demonstrates that a driver's mental state or emotions can influence on how they drive. We build upon this base of traffic psychology and investigate if we can use affective feedback to change an operator's mental state, and therefore change their driving behavior.

The use of social behaviors and strategies follows an established tradition in social robotics, and human-computer interaction in general [31]. For autonomous robots, the use of social behaviors has been shown to influence group communication dynamics [32], [33], dissuade people from performing actions [34], encourage lying to authorities [35], or change how people talk [36]. We see these examples as demonstrating an opportunity to have robots use social phenomena to change and affect interactions and people's behaviors with them [37]. Social behaviors have further been used to communicate robot state (e.g., [10], [11], [38], [39]), discussed above. We extend and combine these strategies using social behaviors *in the teleoperation interface* to communicate state and simultaneously influence the teleoperator themselves.

Social feedback in vehicle driving situations has been shown to be beneficial (e.g., in car interfaces [40]). However, the design of such interfaces is non-trivial, and may be distracting [41] and increase cognitive load [42], [43]. Our design aims to explore emotional displays as a social feedback mechanism, while also exploring how the effects may change teleoperation behaviors.

Social signals and teleoperation are often studied together in the context of telepresence. Telepresence research tries to design robots and robot interfaces that are used by one person to control a robot and interact with another person socially, where the robot is a proxy (e.g., [44]–[47]). Our work contributes to teleoperation by using social feedback mechanisms in cases where there is no human on the remote end: the social interaction is between the operator and the teleoperation interface.

III. Design: Interactive Teleoperation Agents with Affective Feedback

As a proof-of-concept for using affective feedback in teleoperation interfaces, we designed two interactive agent personalities to influence an operator's mental state and potentially robot driving behavior. The agents monitor teleoperation performance in real time, and based on how well the operator is driving, the agents change their facial expression. To explore this space, we designed two different agents, each with a specific affective feedback and reaction strategy. We note that there is a rich potential for future work in applying more complex and thorough psychological frameworks to agent design; our goal here was rather as an exploratory proof-of-concept.

Our design was heavily inspired by the video game *DOOM* (id Software, 1993), where the face of the player's avatar was displayed at the bottom of the screen and reacted emotionally to the avatar's state and events in the environment.

A. Design Strategy: affective feedback

Our approach to influencing an operator is to leverage affective feedback by showing them an emotional reaction to their driving. Previous work has found that when a person sees someone experience an emotion, the viewer may experience a similar emotion (becoming happy when someone around you is happy), often an automatic or reflexive response [7]. Alternatively, if the operator develops empathy for the agent they may react by trying to support the agent [48], [49].

Our goal is to use affective feedback to induce an emotional response in the operator. We do this for the purpose of shaping driving behavior and teleoperation experience. Our exploration concept is that positive emotions will influence behaviors via positive and negative reinforcement: the happy face may make the operator feel happy as well, providing positive reinforcement for the driving behavior at that moment. Conversely, we expect our affective feedback will create negative emotions in the operator if the agent reacts negatively. We expect this to provide negative reinforcement and dissuade the operator from taking similar actions in the future.

With these two ideas in mind, we designed respective interactive agent personas with different affective feedback strategies: an "anxious" and a "daredevil" agent.

B. Personas for affective feedback

Both personas are based on the same principle of trying to encourage certain behaviors with positive emotions and discourage others with negative emotions. Specifically, our agents encourage or dissuade behaviors based on teleoperation danger, such as collisions with obstacles, or driving too quickly. Thus, the reactions act as a social interface that conveys safety information to the operator.

Anxious persona: if an operator drives more dangerously, the agent would become more upset or frightened. Conversely, if the operator drove safely, the agent would become happier. This was to encourage safe driving with happy reactions and dissuade less safe driving.

Daredevil persona: the agent displays an increasingly bored and contemptful face if the operator drives safely but becomes excited if driven dangerously. We expected this persona to promote dangerous driving by providing positive affective feedback when the operator drives dangerously. Further, the negative reactions to safe driving may discourage safe behavior. This was designed to explore if a badly designed persona could possibly promote dangerous behavior.

The daredevil and anxious personas both build on the same approach of leveraging social feedback to change teleoperation behavior, with the different personas helping to explore our strategy.

C. Measuring teleoperation safety

This initial proof of concept uses collisions per minute and robot velocity as coarse measures of driving safety. Collisions are a direct sign of mistakes during operation. Velocity is a measure of safety as, in general, driving very quickly is more dangerous: faster speeds give operators less time to react and not collide with people, expensive equipment, or tumble over a ledge. We acknowledge that a very skilled driver may be able to drive quickly without causing collisions, but they are still subject to these increasing constraints to reaction time and may still make a real (or in our case, virtual) passenger nervous. We concede that our choice of these two measures is a limited representation of safety – reckless acceleration, near misses, and other factors may all contribute to long-term safety. It serves, however, as a sufficient and consistent mechanism for our exploration.

D. Design Implementation

We designed our interactive agent to be easily visible but to not be too distracting. This was done by placing the agent on-screen, overlapping the teleoperation video in a salient location while not covering up a typically important area (Fig. 1b). Further, to provide an illusion of activity for the agent and draw attention [50], we had the agent update its expression twice a second.

Calculating Safety

In order to define how our agents reacted to teleoperation, we had to define what number of collisions per minute and speeds were considered unsafe or safe. We ran pilot experiments to calibrate this, specifically tuning the change in velocity or collisions per minute needed to change the reactions of our personas. Our goal was to find thresholds such that the agents provided noticeable visual and emotional feedback for both the operator's initial driving, and after any changes they may make to their driving in response to the social feedback. Thus, our thresholds are specific to our environment.

We calculated an independent safety rating for both collisions per minute and driving speed, resulting in a value that ranged from most safe to least safe. For collisions, we maintained a running "collisions per minute" total, which summed collisions occurring in the last minute. We used a linear weighted sum to make the agent's changing reaction smoother as older collisions became less relevant: each collision was weighted by how much of a minute had passed since the collision occurred. For example, a collision that was 30 seconds old would contribute to the safety rating as half a collision. Collisions were measured automatically by combining data from the robot's inertial measurement unit and the joystick used to drive the robot.

Velocity-based safety was calculated based on the average velocity over the last minute. We defined "not safe" driving to be anything over a threshold speed (25% of robot max speed). Excess velocity after this threshold was then used to determine the safety rating. As discussed earlier, we did not want to react to maximum speed driving with no collisions as completely unsafe. Thus, max velocity safe driving (no collisions) would only progress the personas to a middle safety state (Fig. 2, neutral).

The final safety rating was the least safe of the two measures, collisions per minute and velocity, recalculated each frame.

Selecting a Reaction

Our interface maps the safety rating, ranging from a minimum safety rating to a maximum rating, to a reaction (Fig. 2). We first ordered the persona's expressions from least safe to safe. Our safety rating is then used as an index in between these expressions; for example, a safety rating of 50% of the maximum safety rating will pick a neutral expression (half way between not safe and very safe expressions). A safety rating of 75% would pick a slightly smiling face, in the case of the anxious person (Fig. 2, top).

Our expressions are taken from video data of people making pre-defined emotional reactions starting from a neutral expression [51], [52]. The personas are formed by reversing the "not safe" emotion video to start from an emotion and end with a neutral expression. We can then transition to the "safe" emotion video by moving between the neutral expressions in both videos.

Thus, each expression in our dataset is a frame in this linked video – a video of an unsafe reaction, transitioning to a neutral reaction, transitioning to a safe reaction. Our safety index is mapped to a frame in this video, which is displayed in our interface (e.g. Fig. 1b). As the safety rating changes, we simply display new frames from the video, providing a smooth emotion transition. If the safety rating stays the same, a nearby, similar frame of video is used to show small movement in the agent, such as slightly moving the corners of their lips or eyes. This creates an illusion of activity and livelihood to the agent, and may

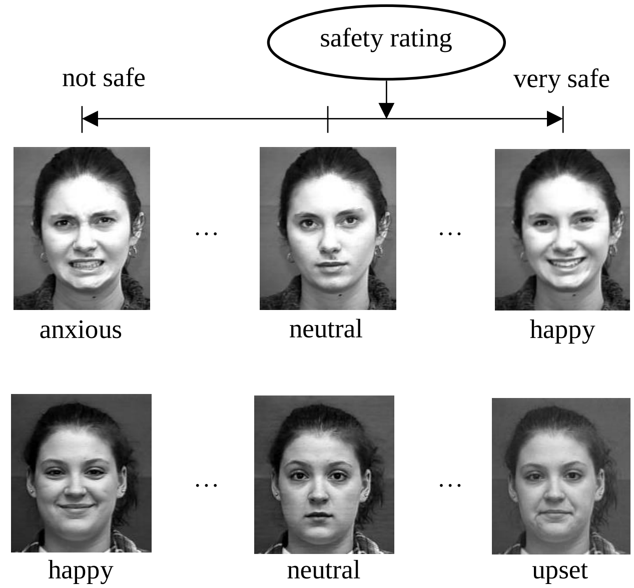

Fig. 2. The range of expressions, mapped from not safe to very safe driving behavior. The real-time driving safety rating indexes into a collection of faces displaying emotion. Top row: *anxious* persona, bottom row: *daredevil* persona.

draw attention to the agent itself [3]. Our emotion video data is from the Extended Cohn-Kanade emotional face dataset to pick our faces [51], [52]. In our anxious persona, we combined "fear" and "disgust" for negative reactions and used "happy" for positive reactions. In our daredevil persona we combined "contempt" and "disgust" for negative reactions and used "happy" for positive reactions.

IV. EXPERIMENT

Our experiment's goal was to investigate the effects of our affective feedback interfaces for teleoperation on the operator's perception of the robot and their driving behavior. To do so, we created a driving task: an obstacle course that would test a participant's ability to control the robot. Participants drove the robot through an obstacle course with the two interactive agents and a base case, and, in each trial, we measured their driving performance, their perceptions of the robot, and their driving using self-report measures.

A. Task

Participants were tasked with remotely driving a robot around an obstacle course. The course consisted of a grid of obstacles and a series of arrows that had to be followed, with each arrow indicating a 90 degree turn around a corner (heavily inspired from previous work [22]). Participants would drive 3 laps around the course, with the first lap being treated as a practice run. We instructed participants to drive as fast as they felt comfortable, while trying to avoid any collisions with obstacles along the course.

We designed three similar obstacle courses for the within-participants study; while the obstacles did not move between trials, the arrows leading them through the course did change. Each course had similar length and number of turns to maintain difficulty across conditions. Further, courses were designed to have a mix of straight sections and sequences of turns to test different driving scenarios.

B. Manipulations

We tested three conditions. The two interactive agents, *anxious* and *daredevil* personas, and a numeric-display base case. We struggled to develop a base case, as our first inclination was to simply have an interface with no feedback. However, this would compare two things: availability of driving feedback, and, emotional encoding. By including the numeric case, we can keep the feedback only without the affect. This base case displayed the same information encoded in our personas but had no social or emotional element (Fig. 3). Each persona started the condition showing the "very safe" reaction. This allowed us to test whether just the information alone could influence an operator's driving in comparison to the social encoding.

Our experiment used a within-participant design; each participant used all interfaces: anxious persona, daredevil persona, and the baseline. Condition order was fully counterbalanced across participants, while course order was fixed for all participants.

C. Measures

Before the experiment, we administered a demographics questionnaire that recorded and age and gender. We further inquired about any experience they have for activities similar to robot teleoperation: experience playing video games, experience driving vehicles, experience with remote control robots (quad copters, RC cars, etc.), and participation in any other robot experiments.

In each condition we recorded the time it took to complete the task and number of collisions. During the experiment, we also logged robot velocity and the current safety rating of the participant's driving. The robot's movement data was recorded as a potential way to measure changes in operation.

Fig. 3. Our baseline interface simply displayed the safety information without social or emotional cues. The text reads: "collisions/min: 2.8 velocity: 68.6%"

To understand changes in self-reported workload, we administered the NASA TLX questionnaire [53]. Further, we measured the operator's emotional state on a common two dimensional emotion model (valence and arousal [54]), with the Self-Assessment Manikin instrument (7-point variant, from -3 to +3) [55]. To measure changes in perception of the robot's operation, we additionally asked participants to rate the robot's overall safety for driving, and informativeness of the safety indicator interface. The post-condition questions included free-form feedback areas for participants to give positive, negative, or other feedback that they felt was appropriate.

In our pilot studies, we noticed participants did not pay attention to the safety interfaces (including the base case), perhaps due to the study being in a safe lab environment. To encourage operators to pay attention to the safety information, we created a distractor question about the information displayed: we ask operators to choose "the face shown most often while you drive," or "the average velocity you thought you were closest to most often." Then, we show a range of five faces used by the agent during the condition, spread from negative to positive emotions. For the baseline, five percentages of max velocity, spaced from 20% to 100% are shown. This question was not for analysis, but to make participants pay attention.

After the experiment, we asked participants to rank each interface for preference. There were also optional short answer blocks for comments, similar to those described above in the post-condition questionnaire. Finally, we administered a questionnaire from prior work that measures susceptibility to emotional responses when exposed to different emotions, from [56], which we thought may help control for variance in our observations.

D. Procedure

Participants are welcomed and told we will be exploring ways to convey safety information to robot operators and investigating how that may affect how safely they drive the robot. We explicitly tell participants we are using collision information and robot speed to gauge driving safety, and that this information will be displayed via a summary as a facial expression. We do not, however, state which expression correlates to what driving safety level. The consent form and demographics questionnaire are filled out at this point.

Each persona is first introduced and explained using a paper representation, with multiple expressions shown (similar to Fig. 2, but with positive emotions aligned to the same side). We additionally introduce the baseline system (Fig. 3), and explain that the information it displays is the exact same information used by the system's algorithm to decide what face is shown (words like system and algorithm are used, emphasizing the mechanical nature of our interface, and not implying our agent is intelligent).

The participants are instructed that their task is to drive through the obstacle course as fast as they feel comfortable while trying to avoid all obstacles. After course instructions and controls are demonstrated, participants are given one lap to practice. Afterwards, they drive two laps with the same agent and course, during which data is recorded. If necessary, after the practice lap, obstacles are replaced in case they were pushed around, and the program is restarted to reset the agents to a "very safe" state. From pilots, we found each lap of our courses took around one to five minutes a lap, depending on participant skill. We found participants took around two to three extra minutes on their practice laps as well, resulting in roughly 15 to 54 minutes of driving per person.

Before the two laps where data is recorded in each condition, we explain the distractor question to participants, so they know to pay attention to the agent. After the laps are complete, the distractor question and other post-condition questionnaires are administered. The next obstacle course is prepared, the new on-screen interface (interactive agent or baseline) is explained, and the participant is given a practice lap before continuing.

At the end of the experiment, participants are given the post-experiment questionnaire (interface ranking, final comments), and brought to see the course and robot in person. The details of the experiment are explained, as well as why we were purposefully vague on how the agents each conveyed the safety information. After any questions were answered, the experiment was over.

E. Implementation

Our robot was a Clearpath Jackal robot running ROS Indigo. It was limited to 50% of its maximum forwards and backwards speed, and 75% of its maximum turning speed as pilot testing showed our robot moved too quickly in our smaller environment. A PointGrey Flea3 camera was mounted near the front of the robot such that the robot itself was not in the view of the camera. The camera was run in 640x480 resolution (Fig. 1b) at 45 frames per second over the institution's Wi-Fi network. The data handling and networking was handled through multi-threaded python code.

Participants were seated at a desk and allowed to adjust the setup to be comfortable. They used a 4K 27-inch monitor, with the interface maximized (black bars were used for letterboxing). They controlled the robot with a joystick (Microsoft Sidewinder USB) on the desk in front of them. The client-side was programmed in C#.

In our pilot studies we found that the robot was able to move our obstacles easily, hindering the study. To make the obstacles more stable, they were each weighted with 14 KG of weights, placed on rubber friction mats, which in turn were placed on carpet tape stuck to the ground. With this much resistance, the robot could not easily push obstacles out of the way: operators needed to navigate the obstacle course correctly. To further emphasize collisions, our system would make the whole screen flash red briefly (1/3 of a second) when a collision was detected.

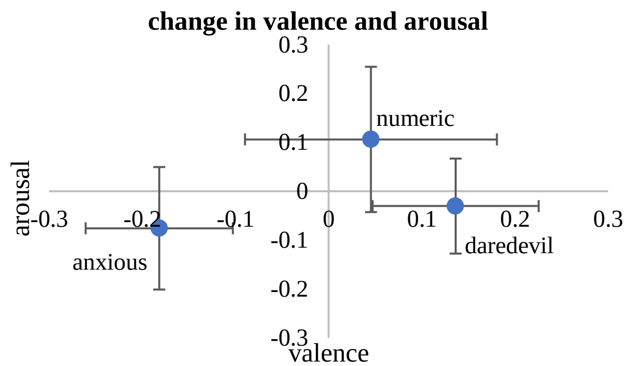

Fig. 4. The reported emotions of operators after using each interface. Anxious interface appeared to lower valence more than the numeric case, while daredevil had lower arousal than the numeric (from contrasts). Grand mean differences are p<.05. Error bars show standard error.

F. Analysis

We investigate the two components of our affective feedback strategy: a) how the agent behavior impacted operator mental state (if we induced an emotional response), and b) how this impacted the operator's driving behavior and teleoperation experience. The emotion analysis included the five emotion-susceptibility questionnaire subscales as covariates, to control for how not everyone is affected by displays of emotion equally.

For performance measures, we analyzed collisions over time (number of collisions divided by completion time, in minutes), perceived workload (TLX sum and its subscales), our perceived safety and informative scales, and the safety rating calculated by our system.

Results

We recruited 23 participants by advertising with posters around our local university area. One participant did not complete the experiment, resulting in 22 participants (mean age of 28, standard deviation of 10.7 years; 10 female).

To understand how our interfaces may have changed operators emotionally, we ran a repeated-measures ANOVA on reported measurements for valence and arousal changes, with the five emotion-susceptibility questionnaire subscales as covariates. We found a statistical effect of the interface on self-reported valence (a measure of pleasure or displeasure, $F_{2,32}=4.1$, $p<.03$, $\eta^2=.20$), and arousal (a measure of activation, or sleepiness, $F_{2,32}=3.4$, $p<.05$, $\eta^2=.18$). Post-hocs with Bonferroni correction found the daredevil agent produced higher self-reported valence than the anxious agent ($p<.02$, mean difference=-.32 points, 95% CI [-.59, -.05]). Other pairwise comparisons were non-significant.

Marginal means showed the daredevil case had the highest valence (mean=0.36, 95% CI [-0.07, 0.80]), followed by the numeric case (mean=0.27, 95% CI [-0.24, 0.79], with the anxious agent (mean=.05, 95% CI= [-0.44, 0.53]), having the lowest valence. For arousal, we found the numeric interface had the highest (mean=-0.68, 95% CI [-1.34, -0.02]), followed by the daredevil interface (mean=-0.82, 95% CI [-1.35, -0.29]), with the anxious case have the lowest arousal (mean=-0.86, 95% CI [-1.30, -0.42]). See Fig. 4 (note for legibility, we have enlarged the graph, but the scale ranged from -3 to +3).

There was an interaction effect between the subscale on susceptibility to happy emotions and the interface ($F_{2,32}$=4.4, p=.02, η^2=.22). We present the graph of the interaction (Fig. 7) but note there were few participants for each valence rating, so we caution drawing conclusions from it.

We performed repeated measures ANOVAs on the performance measures listed above. The effect of the interface on collisions over time (CPM) was non-significant ($F_{2,42}$=2.6, p=.085, η^2=.11). Marginal means showed the numeric case had the most collisions (mean=1.9 CPM, 95% CI [1.6 CPM, 2.2 CPM], followed by the anxious agent (mean=1.8 CPM, 95% CI [1.5 CPM, 2.2 CPM]), with the daredevil agent interface having the fewest (mean=1.6 CPM, 95% CI [1.3 CPM, 1.9 CPM]) – see Fig. 5.

We wished to test if the interface may have improved operation over time, as exposure to the reactions potentially affected driving behavior as time passed. We ran a 2x2 ANOVA (interface versus time), with sample points at 10% intervals throughout the experiment. This was not found to be significant (p>.05) – see Fig. 6.

The agent's reactions may have been used by operators to inform themselves of their performance, but we found no statistical effect of the interface on perceived performance ($F_{2,42}$=2.7, p=.08, η^2=.11) – note the TLX performance scale is reverse-coded and higher scores mean worse perceived performance. Marginal means showed the daredevil case made participants feel they performed the best (mean=7.8 points, 95% CI [6.1 points, 9.5 points], followed by the numeric case (mean=8.8 points, 95% CI [6.9 points, 10.7 points]). The anxious agent interface made had participants feel they performed worst (mean=9.4 points, 95% CI [7.5 points, 11.3 points]) – see Fig. 8.

The perceived informativeness of the interface had a statistical difference ($F_{2,42}$=3.9, p=.03, η^2=.16) Marginal means showed the numeric case was perceived to be the most informative (mean=14.9 points, 95% CI [12.8 points, 16.9 points], followed by the anxious agent interface (mean=13.9 points, 95% CI [11.7 points, 16 points]). The daredevil agent interface was considered the least informative (mean=12.8 points, 95% CI [10.3 points, 15.3 points]) – see Fig. 9.

All other tests and interactions were found to be non-significant.

V. DISCUSSION

Our results found differences in self-reported valence and emotion after using our interface, implying that their emotions did change somewhat for each interface. We found inconclusive evidence for the numeric case to have a higher collision per minute rating than the social interfaces, and that collision rates may be stable over time, though future study is required to confirm this. Although we did not detect a statistical difference, our results indicating potential differences merit further inquiry. Daredevil had the lowest average collisions per minute and was perceived by operators as enabling the better performance of our three interfaces. The anxious interface had fewer mean collisions per minute than the numeric case, but was seen as having worse performance. The numeric interface was the most informative, followed by the anxious and then daredevil interfaces.

The changes in valence and arousal demonstrate that an on-screen agent using affective feedback of safety ratings can change an operator's mental state. Interestingly, when inspecting average safety scores, we found that people, on average, drove in a way that our system rated as unsafe. Thus, people would have seen primarily a negative reaction from the anxious agent, and a happier face for the daredevil agent. This aligns with our background theory and results:

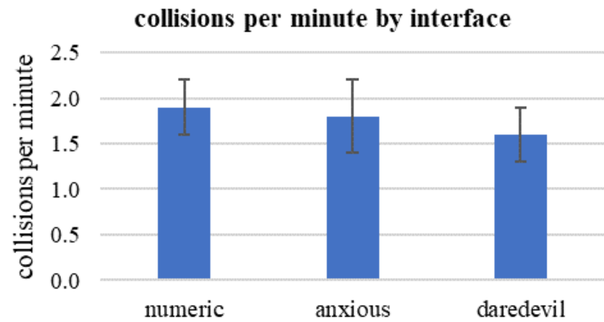

Fig. 5. The average collision density during operation depending on which interface operators saw. Means are not significant (p=.085). Error bars show 95% CI.

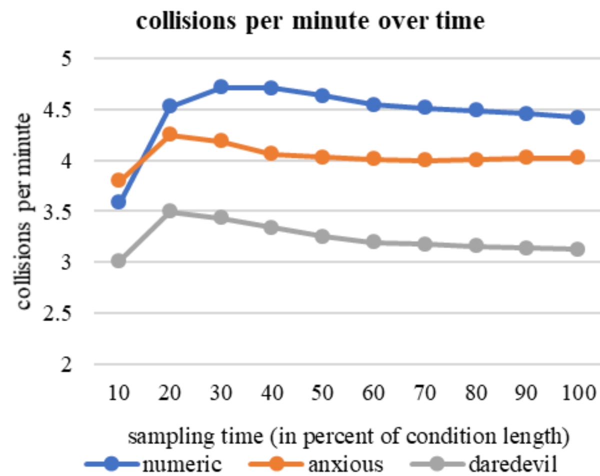

Fig. 6. Collisions per minute sampled at 10% intervals through the study. High variance in our sample means we could not detect a difference (p>.05).

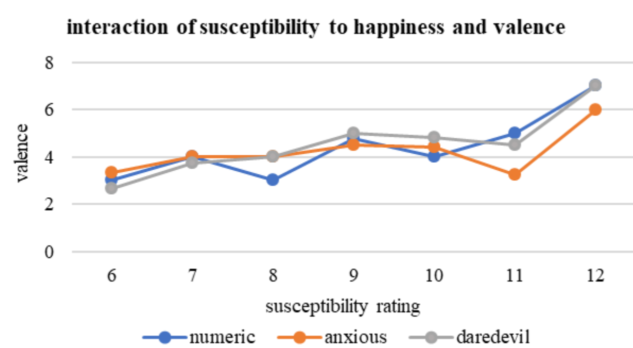

Fig. 7. The interaction of an operator's susceptibility to displays of happiness, measured by questionnaire, and valence, by interface. The interaction is significant (p<.05)

the anxious interface (a lower valence emotion than happiness [54]) was reported as making participants feel lower valence overall, and happiness (a higher valence emotion [54]) had a higher valence. Thus, we can see the expected emotion divide (happy, sad) between daredevil and anxious, acting as a manipulation check that viewing emotions in a teleoperation interface can influence the operator's emotions to become more similar to the displayed affective behavior. While the differences were small, we note that the interaction overall

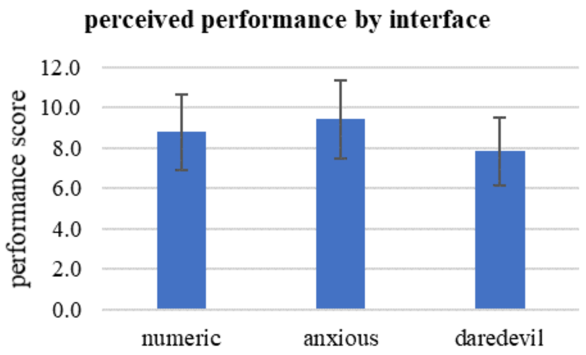

Fig. 8. Self-reflection performance values by operators were not significant (p=.08). Performance is reverse-coded (higher means worse perceived performance).

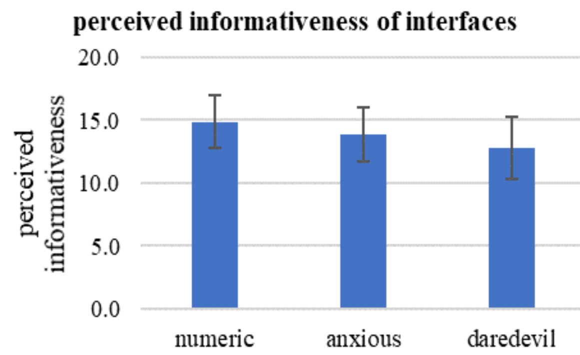

Fig. 9 The perceived informativeness of the interfaces was different (p<.5). Error bars show 95% CI. Interestingly, the difference with the numeric case was slight, despite very different visualizations.

was very short. Small differences over time, however, may amount to a longer-term effect, but more research is needed to confirm this.

Our theory that positive emotions would encourage the behavior at the time of the affective feedback was not supported by our data. It is possible the emotional response itself was not strong enough for this effect to take place. Another possibility is that the daredevil persona helped people relax; when colliding, the reaction on the face was happiness, which may have reassured the participant. If they saw the anxious persona look unhappy and experienced our observed negative valence shift, instead of discouraging the behavior, the feedback may have made them tense up and perform worse. This may also explain why self-reported performance was higher for the daredevil persona: the positive reaction upon mistakes made participants think they were not doing poorly. Certainly, the intricacies of how participants reacted to the emotion needs further research for clarification. Further, this highlights the importance of a more rigorous model for creating personas, which would enable us to more concretely and specifically reflect on components of the agent's reaction.

Even over a short period of operation, such as 30 minutes, a difference of 0.3 collisions per minute (the same difference between our numeric and daredevil case) results in an extra 10 collisions. Further, as per Fig. 6, the average difference in collisions per minute between interfaces may be stable over time. Large variances, however, stop us from being able to reach strong conclusions, and require further research. It is interesting that the numeric interface appears to have higher collisions per minute. This may be due to the mental work needed to read and understand numeric data, which may take more attention away from actual operation. However, we stress these collision results are not statistically significant, and further study is needed to confirm if our measures are correct.

The daredevil interface was seen as least informative, which may be due to it being unintuitive: after seeing a more positive face after a collision, operators may have thought the system was not working properly. However, all three interfaces were ranked similarly (Fig. 9), which may suggest that social interfaces for communicating information may be feasible when the operator does not need a granular understanding of data, such as in our case.

While we motivated our design with existing social psychology theory, social interfaces have sometimes struggled in industry. Microsoft's Clippy is one example: critiques of Clippy point out that Clippy breaks social rules, such as offering help when it is not asked for and not remembering people's decisions [57]. Others have argued virtual companions should be more agreeable (such as our daredevil showing happy faces during bad driving), or offer alternative solutions to a difficult task [58], [59]. Our agents were only reactive and did not try to provide advice to users; this may be why we did not witness similar negative feedback to our agents. It is possible that virtual agents helping in teleoperation may have different social rules applied to them and is an important avenue for future work.

VI. LIMITATIONS AND FUTURE WORK

Our results raised many questions for future work. Overall, we aimed for external validity and used an algorithm that could be applied to robots right now, but that made it difficult to evaluate the nuances of the social interfaces. For example, a future study could look at if the agent even needs to properly react to current driving behavior: it could always look annoyed, or happy. This would remove the variable in our studies where drivers of different skill levels would see, on average, different agent reactions. If certain agent reactions would have a stronger effect than others, we would have difficulty measuring those effects as our operators had different levels of exposure to each reaction.

The agents were described primarily as a tool, or algorithm. It is possible that this reduced the anthropomorphism effect and reduced the impact of the agent's reactions. If operators thought the agent was intelligent, it is possible they would react to the agent in a more social way. Further, the agent could be presented in multiple different ways: as a boss, as a coworker, as the robot's intelligence, etc. This change in agency and relationship with the operator could further affect their reactions to the agent's displayed emotion.

The emotion susceptibility questionnaire we used had five subscales, which proved difficult to use with a smaller participant pool. With our participant numbers, adding all scales as covariates may lead to overfitting for our model. Due to the direct link between emotion susceptibility and viewer response in the literature, we believed including the covariates for our emotion data (valence and arousal) was necessary. However, we opted to use a simpler model in our performance analysis for fear of overfitting.

We witnessed a large amount of variance between participants. This may be due to our course being overly difficult, such as having little room for the robot to make turns and operators having difficulty visualizing the robot in the remote area. Anecdotally, we witnessed operators who performed well, but got stuck in difficult situations on occasion, resulting in numerous collisions during a single event, perhaps increasing the variance in our results. While we pilot tested extensively to calibrate our agents, we still believe that the course difficulty may have confounded our results. We recommend future work carefully consider and calibrate the difficulty of their study.

Our choice of baseline may contribute to our results. We opted to design our baseline to have information parity with our affective feedback agents: all interfaces, on some level, presented collision and velocity information. However, by displaying safety information numerically, we likely increased the mental processing needed to understand the presented information as compared to the affective feedback from our agents. To reduce this, we could have a baseline with no information displayed, or just a neutral face displayed, or to use simple text labels for the emotional state (e.g., "safe", "unsafe", "very unsafe", etc.). Thus, our baseline is not a truly neutral control condition, but enables us to compare social to non-social interfaces without the confound of difference in available information.

One major limitation we believe is the presence of the researcher in the room while the participant was piloting the robot. As the researcher was an authority figure and a stranger, the operator may have suppressed their reactions to appear more professional and under control to the researcher. The researcher's own subtle and subconscious body language may have provided a stronger social signal to the operator than the agent itself. Thus, we recommend removing the researcher from the room in future studies.

VII. CONCLUSION

We presented a proof-of-concept for using affective agents to shape teleoperation experience and highlights the potential for using social HRI techniques between the teleoperated robot and the operator. Our results highlight using social interfaces designed to leverage affective feedback can indeed change a teleoperator's mental state, and its design impacts how effective they felt the feedback was for communicating state. Further work is needed to leverage change in emotional state to improve actual teleoperation performance. Our results pave the way for more research into applying other psychological and social phenomenon to teleoperation.

VIII. REFERENCES

[1] H. A. Yanco and J. Drury, "'Where am I?' Acquiring situation awareness using a remote robot platform," in *IEEE Systems, Man and Cybernetics*, 2004, vol. 3, pp. 2835–2840.

[2] J. Y. C. Chen, M. J. Barnes, and M. Harper-Sciarini, "Supervisory control of multiple robots: Human-performance issues and user-interface design," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 41, no. 4, pp. 435–454, 2011.

[3] D. J. Rea, S. H. Seo, N. Bruce, and J. E. Young, "Movers, Shakers, and Those Who Stand Still: Visual Attention-grabbing Techniques in Robot Teleoperation," in *Human-Robot Interaction*, 2017, pp. 398–407.

[4] S. Giese, D. Carr, and J. Chahl, "Implications for unmanned systems research of military UAV mishap statistics," *IEEE Intell. Veh. Symp. Proc.*, no. Iv, pp. 1191–1196, 2013.

[5] K. W. Williams, "A Summary of Unmanned Aircraft Accident / Incident Data: Human Factors Implications," *FAA Civ. Aerosp. Med. Inst.*, no. December 2004, p. 18, 2004.

[6] G. Schoenewolf, "Emotional Contagion: Behavioral Induction in Individuals and Groups.," *Mod. Psychoanal.*, vol. 15, no. 1, pp. 49–61, 1990.

[7] E. Hatfield, J. Cacioppo, and R. Rapson, "Emotional contagion," *Current Directions in Psychological Science*, vol. 2, no. 3. pp. 96–99, 1994.

[8] S. G. Barsade, "The Ripple Effect: Emotional Contagion and Its Influence on Group Behavior," *Adm. Sci. Q.*, vol. 47, no. 4, p. 644, 2002.

[9] J. E. Young, R. Hawkins, E. Sharlin, and T. Igarashi, "Toward acceptable domestic robots: Applying insights from social psychology," *Int. J. Soc. Robot.*, vol. 1, no. 1, pp. 95–108, 2009.

[10] M. Sharma, D. Hildebrandt, G. Newman, J. E. Young, and R. Eskicioglu, "Communicating Affect via Flight Path," in *Human Robot Interaction 2013*, 2013, pp. 293–300.

[11] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin, "Effects of Nonverbal Communication on Efficiency and Robustness of Human-Robot Teamwork.pdf," *IEEE Int. Conf. Intell. Robot. Syst.*, 2005.

[12] A. Singh and J. E. Young, "A dog tail for utility robots: Exploring affective properties of tail movement," *Lect. Notes Comput. Sci.*, vol. 8118 LNCS, no. PART 2, pp. 403–419, 2013.

[13] S. H. S. H. Seo, D. J. D. J. Rea, J. Wiebe, and J. E. J. E. Young, "Monocle: Interactive Detail-in-Context Using Two Pan-and-Tilt Cameras to Improve Teleoperation Effectiveness," in *RO-MAN*, 2017, vol. 2017–Janua.

[14] J. Y. C. Chen, E. C. Haas, and M. J. Barnes, "Human Performance Issues and User Interface Design for Teleoperated Robots," *IEEE Trans. Syst. Man Cybern. Part C (Applications Rev.*, vol. 37, no. 6, pp. 1231–1245, 2007.

[15] A. Singh, S. H. Seo, Y. Hashish, M. Nakane, J. E. Young, and A. Bunt, "An interface for remote robotic manipulator control that reduces task load and fatigue," *RO-MAN*, pp. 738–743, Aug. 2013.

[16] a Leeper, K. Hsiao, M. Ciocarlie, L. Takayama, and D. Gossow, "Strategies for human-in-the-loop robotic grasping," *Human-Robot Interact. (HRI), 2012 7th ACM/IEEE Int. Conf.*, pp. 1–8, 2012.

[17] C. W. Nielsen, M. A. Goodrich, and R. W. Ricks, "Ecological interfaces for improving mobile robot teleoperation," *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 927–941, 2007.

[18] D. J. Rea, M. R. Hanzaki, N. Bruce, and J. E. Young, "Tortoise and the Hare Robot Slow and steady almost wins the race , but finishes more safely," in *Robot and Human Interactive Communication (RO-MAN)*, 2017, pp. 1–6.

[19] J. E. Young, Y. Kamiyama, J. Reichenbach, T. Igarashi, and E. Sharlin, "How to walk a robot: A dog-leash human-robot interface," in *2011 RO-MAN*, 2011, pp. 376–382.

[20] D. Sakamoto, K. Honda, M. Inami, and T. Igarashi, "Sketch and run: a stroke-based interface for home robots," *Conf. Hum. Factors Comput. Syst.*, pp. 197–200, 2009.

[21] S. Radmard, A. Moon, and E. A. Croft, "Interface Design and Usability Analysis for a Robotic Telepresence Platform," in *RO-MAN*, 2015, p. 6.

[22] D. J. Rea and J. E. Young, "It's All in Your Head," in *ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, 2018, pp. 32–40.

[23] A. Hacinecipoglu, E. I. Konukseven, and A. B. Koku, "Evaluation of haptic feedback cues on vehicle teleoperation performance in an obstacle avoidance scenario," in *2013 World Haptics Conference, WHC 2013*, 2013.

[24] J. A. Groeger and J. A. Rothengatter, "Traffic psychology and behaviour," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 1, no. 1, pp. 1–9, 1998.

[25] J. A. Groeger, "Trafficking in cognition: Applying cognitive psychology to driving," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 5, no. 4, pp. 235–248, 2002.

[26] C. Eyssartier, S. Meineri, and N. Gueguen, "Motorcyclists'

intention to exceed the speed limit on a 90km/h road: Effect of the type of motorcycles," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 45, pp. 183–193, 2017.

[27] J. A. Michon, "A critical view of driver behavior models: what do we know, what should we do?," *Hum. Behav. traffic Saf.*, pp. 485–520, 1985.

[28] L. Precht, A. Keinath, and J. F. Krems, "Effects of driving anger on driver behavior – Results from naturalistic driving data," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 45, pp. 75–92, 2017.

[29] M. Blommer, R. Curry, R. Swaminathan, L. Tijerina, W. Talamonti, and D. Kochhar, "Driver brake vs. steer response to sudden forward collision scenario in manual and automated driving modes," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 45, pp. 93–101, 2017.

[30] R. C. McIlroy, N. A. Stanton, and L. Godwin, "Good vibrations: Using a haptic accelerator pedal to encourage eco-driving," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 46, pp. 34–46, 2017.

[31] J. R. Lee and C. I. Nass, "Trust in Computers: The Computers-Are-Social-Actors (CASA) Paradigm and Trustworthiness Perception in Human-Computer Communication," in *Trust and Technology in a Ubiquitous Modern Environment: Theoretical and Methodological Perspectives*, IGI Global, 2010, pp. 1–15.

[32] M. F. Jung, N. Martelaro, and P. J. Hinds, "Using Robots to Moderate Team Conflict," in *Human-Robot Interaction*, 2015, pp. 229–236.

[33] D. Sakamoto and T. Ono, "Sociality of robots: do robots construct or collapse human relations?," in *Human-robot interaction*, 2006, p. 355.

[34] G. Briggs and M. Scheutz, "How Robots Can Affect Human Behavior: Investigating the Effects of Robotic Displays of Protest and Distress," *Int. J. Soc. Robot.*, vol. 6, no. 3, pp. 343–355, 2014.

[35] P. H. Kahn, T. Kanda, H. Ishiguro, B. T. Gill, S. Shen, H. E. Gary, and J. H. Ruckert, "Will People Keep the Secret of a Humanoid Robot?," in *Human-Robot Interaction -*, 2015, pp. 173–180.

[36] J. Brandstetter, C. Beckner, E. B. Sandoval, and C. Bartneck, "Persistent Lexical Entrainment in HRI," in *Human-Robot Interaction*, 2017, pp. 63–72.

[37] B. Postnikoff and I. Goldberg, "Robot Social Engineering," in *Companion of Human-Robot Interaction*, 2018, pp. 313–314.

[38] A. Singh and J. E. Young, "Animal-inspired human-robot interaction: A robotic tail for communicating state," in *Human-Robot Interaction*, 2012, pp. 237–238.

[39] J. E. Young, E. Sharlin, and T. Igarashi, "Teaching Robots Style: Designing and Evaluating Style-by-Demonstration for Interactive Robotic Locomotion," *Human–Computer Interact.*, vol. 28, no. 5, pp. 379–416, Sep. 2013.

[40] G. Leshed, T. Velden, O. Rieger, B. Kot, and P. Sengers, "In-car gps navigation: engagement with and disengagement from the environment," *Proc. SIGCHI Conf. Hum. factors Comput. Syst. (CHI '08)*, pp. 1675–1684, 2008.

[41] R. Srinivasan and P. P. Jovanis, "Effect of Selected In-Vehicle Route Guidance Systems on Driver Reaction Times," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 39, no. 2, pp. 200–215, Jun. 1997.

[42] M. Blanco, W. J. Biever, J. P. Gallagher, and T. A. Dingus, "The impact of secondary task cognitive processing demand on driving performance," *Accid. Anal. Prev.*, vol. 38, no. 5, pp. 895–906, Sep. 2006.

[43] J. L. Drury, J. Scholtz, and H. a. Yanco, "Awareness in human-robot interactions," *IEEE Int. Conf. Syst. Man Cybern.*, vol. 1, no. October, 2003.

[44] M. Rueben, F. J. Bernieri, C. M. Grimm, and W. D. Smart, "Framing Effects on Privacy Concerns about a Home Telepresence Robot," in *Human-Robot Interaction*, 2017, pp. 435–444.

[45] A. Kristoffersson, S. Coradeschi, and A. Loutfi, "A review of mobile robotic telepresence," *Adv. Human-Computer Interact.*, vol. 2013, 2013.

[46] K. M. Tsui, A. Norton, D. J. Brooks, E. McCann, M. S. Medvedev, and H. A. Yanco, "Design and development of two generations of semi-autonomous social telepresence robots," in *2013 IEEE Conference on Technologies for Practical Robot Applications (TePRA)*, 2013, pp. 1–6.

[47] K. Tanaka, N. Yamashita, H. Nakanishi, and H. Ishiguro, "Teleoperated or autonomous?: How to produce a robot operator's pseudo presence in HRI," in *Human-Robot Interaction*, 2016, pp. 133–140.

[48] C. Bartneck, M. van der Hoek, O. Mubin, and A. Al Mahmud, "'Daisy, Daisy, give me your answer do!,'" *Proceeding ACM/IEEE Int. Conf. Human-robot Interact. - HRI '07*, no. 2007, p. 217, 2007.

[49] S. H. Seo, D. Geiskkovitch, M. Nakane, C. King, and J. E. Young, "Poor Thing! Would You Feel Sorry for a Simulated Robot?," *Proc. Tenth Annu. ACM/IEEE Int. Conf. Human-Robot Interact. - HRI '15*, pp. 125–132, 2015.

[50] S. L. Franconeri and D. J. Simons, "The dynamic events that capture visual attention: A reply to Abrams and Christ (2005).," *Percept. Psychophys.*, vol. 67, no. 6, pp. 962–6, 2005.

[51] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression," in *Workshop on CVPR for Human Communicative Behavior Analysis*, 2010.

[52] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Automatic Face and Gesture Recognition*, 2000.

[53] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index)" in *Human Mental Workload*, 1988, pp. 139–183.

[54] J. Posner, J. A. Russell, B. S. Peterson, J. A. Russellc, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Dev. Psychopathol.*, vol. 17, no. 3, pp. 715–734, 2005.

[55] J. Morris, "Observations: SAM: the Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response," *J. Advert. Res.*, no. December, 1995.

[56] R. Doherty, "The emotional contagion scale: A measure of individual differences," *J. Nonverbal Behav.*, vol. 21, no. 2, pp. 131–154, 1997.

[57] B. Whitworth, "Polite computing," *Behav. Inf. Technol.*, vol. 24, no. 5, pp. 353–363, 2005.

[58] M. Luria, "Designing Robot Behavior Based on Fictional Sidekick Characters," in *Human-Robot Interaction*, 2018, pp. 307–308.

[59] M. Vázquez, A. Steinfeld, S. E. Hudson, and J. Forlizzi, "Spatial and other social engagement cues in a child-robot interaction," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14*, 2014, pp. 391–398.