Subliminal Priming in Human-Agent Interaction: Can Agents Use Single-Frame Visuals in Video Feeds to Shape User Perceptions?

Elaheh Sanoubari, Denise Y. Geiskkovitch, Diljot S. Garcha, Shahed A. Sabab, Kenny Hong, James E. Young, Andrea Bunt, Pourang Irani
Department of Computer Science, University of Manitoba
{e.sanoubari, d.geiskkovitch, garchads, sabab05, hongk, young, bunt, irani}@cs.umanitoba.ca

ABSTRACT

We investigated interactive agents using subliminal priming – the act of exposing a person to stimuli that they may not consciously notice, but are still processed subliminally in their mind – in an attempt to shape a person's mood and behavior. We present an overview of the psychology of subliminal priming from the perspective of how it applies to human-agent interaction, including a discussion of the potential ethical and practical implications. We further present the results from two exploratory studies (one in-lab, one crowdsourced) that present potential subliminal-priming interfaces. Our results suggest that subliminal priming may impact how participants perceive an agent and how much they enjoy a task, but we failed to find any effect of priming on participant mood or agent persuasiveness. This work aims to raise awareness of the dangers of subliminal methods of priming and contributes to the discussion on the ethics of social agents.

KEYWORDS

Human-agent interaction; persuasive agents; persuasion; priming; subliminal priming; visual priming; trustworthy interaction.

ACM Reference Format

Elaheh Sanoubari, Denise Y. Geiskkovitch, Diljot S. Garcha, Shahed A. Sabab, Kenny Hong, James E. Young, Andrea Bunt, Pourang Irani. 2018. Subliminal Priming in Human-Agent Interaction: Can Agents Use Single-Frame Visuals in Video Feeds to Shape User Perceptions?. In 6th International Conference on Human-Agent Interaction (HAI'18), Dec. 15–18, 2018, Southampton, United Kingdom. ACM, NY, NY, USA, 5 pages. https://doi.org/10.1145/3284432.3284447

1 INTRODUCTION

Interactive agents, whether on-screen or as robots, are being researched for use in many personal contexts, such as serving as personal assistants [40], educators [48], or health coaches [33]. In many cases, an agent's success is measured by its ability to have an effect on a person's behavior or decisions [15]. We investigate how agents can use subliminal cues, messages that are not explicitly noticed by a person but get processed in their brain, to shape one's perception of the agent and alter their mood and behavior.

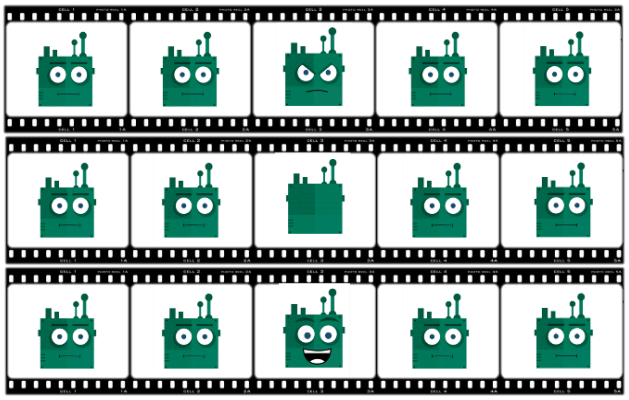

Figure 1. Still-frame sequences from the video of our agent. The video is laced with single-frame emotionally-charged stimuli (angry, neutral and happy) in attempt to shape user mood and perception.

A large body of work in psychology has been dedicated to exploring social persuasion strategies for use during interaction with others (e.g., [54]). Many persuasion techniques involve using explicit social cues, such as smiling or giving a compliment, to try to shape a person's perceptions and how they interact with others. There are also more subtle methods of persuasion, such as making slight changes to one's voice [13], the use of specific colors [25], or even designing the context and environment (e.g., to include angry or calm imagery), in an attempt to shape the interaction. *Subliminal priming* is the use of stimuli that can affect someone, without them being consciously aware of it. We explore if agents can use subliminal priming to shape people's perceptions of the agent, and potentially their behaviors and interactions with it.

Agents, both animated on-screen or physically embodied in robots, draw user attention during interaction; for example, people can be expected to pay attention and look directly at an agent while it is talking, gesturing, or is animated in other ways. We explore how an agent can leverage this; while the agent has a person's attention, it can give subliminal social cues to attempt to change the person's mood and behavior. Specifically, our agent flashes images that are emotionally charged: these images last a

HAI '18, December 15–18, 2018, Southampton, United Kingdom

ACM ISBN 978-1-4503-5953-5/18/12...$15.00
https://doi.org/10.1145/3284432.3284447

single frame (in a 30 frames per second video), embedded in a video feed that the person is focused on (as in Figure 1). This duration (33ms) is well above the threshold (of 16ms) established in psychology literature for subliminal visual priming [66], to ensure that the priming stimuli are processed by the participants.

We base our work on a rich background of subliminal priming in Psychology. However, previous psychological studies investigating the efficacy of subliminal priming have had mixed results, with the impact of priming heavily influenced by the context and specifics of the methodology [58]. We conducted two exploratory experiments to investigate specific subliminal priming techniques for human-agent interaction: a laboratory study using a humanoid robot with affixed LCD display (which flashed happy, angry, or neutral human faces for subliminal priming), and a crowdsourced study using videos of an animated agent (that flashed happy, angry, or neutral agent faces). Our results in both studies suggest an impact of subliminal priming on user perceptions of the agent and task enjoyability, but did not find an effect on participant mood or agent persuasiveness. Further, in both cases we found negative subliminal priming (angry faces) to improve participant mood and opinions of the agent, while positive subliminal priming (happy faces) had the opposite effect. This was surprising and counter to our expectations of positive priming improving participant mood and opinions of the agent. We discuss this and provide several potential avenues for further inquiry.

The results of this work suggest that agents may be able to use subliminal priming, and more work is needed to fully explore the extent of how powerful subliminal priming can be to change a person's mood, perceptions of the agent, or actions. However, even the possibility of an agent having the power to shape users' mood, decisions, and actions, without their explicit knowledge, raises ethical concerns that necessitate further inquiry into agents using subliminal priming. As such, a core contribution of this work is to raise awareness of this potential persuasive method.

2 RELATED WORK

In recent years, research has explored the use of intelligent agents for persuasion and to alter human behaviors in areas such as education [16], health care [11, 12] and energy efficiency [44]. Researchers have investigated manipulating, for example, human-agent proximity, agent gestures, gaze patterns, facial expressions, touching, vocal tone, and vocal expressions, to engineer agent persuasiveness [3, 16, 17]. In these cases, the persuasion is explicit and could be recognized by the user. We extend this work to consider more subtle methods, specifically subliminal priming.

Other work in HAI has explored more generally how agent design can shape attitudes and persuasiveness, for example, based on human-likeness or embodiment of the agent [59], its perceived gender [57], or language use [61], to leverage existing social norms. Such existing work may be more subtly persuasive, for example by leveraging gender stereotypes people may not be explicitly aware of [23]. We build on the success of this work by investigating affective priming stimuli.

Prior work in computer security argues that intelligent agents can use their social and persuasive abilities to perform "social engineering attacks" [17,51]. In a recent study a robot managed to use social tools, such as lying, to convince people to let it into a secure-entrance facility [9]. In this study 87% of the participants who identified the robot as a potential bomb threat still helped it enter the facility. Building on this line of research, we discuss how persuasive agents are already capable of using social tools that introduce challenges to human-agent interaction.

A persuasive agent can use social tools such as asserting authority to pressure users to comply to its requests. In an experiment that studied participants' responses when a robot asked them to perform embarrassing medical procedures [6], more than half of the participants undressed to their underwear at an agent's command and several participants complied to the agent's request when it asked them to measure their rectal temperature. Similar studies provide proof of concept that agents can use social tools to override user judgement of right or wrong. In an experiment studying how participants interact with a faulty robot, more than the majority of the participants followed a robot's request to pour orange juice on a plant [55]. As such, deployment of persuasive agents has raised ethical and practical concerns, leading researchers to discuss the implications of such technologies [47]. We contribute to the on-going research on ethics of persuasive robots by introducing subliminal priming to human-agent interaction.

Further, prior work has moved beyond leveraging social and cultural norms to more intentioned priming aiming to change user behavior and actions. For example, researchers have explored how framing a robot as a social agent versus a machine can change the social behaviors of children towards the robot [64]. Such attempts at priming can improve human-agent interaction. For instance, a teleoperation study that primed participants by misrepresenting a robot's capabilities relating to safety [52], shows that this can result in safer driving behaviors.

Additionally, priming can be used to change user perceptions of the agent. Previous research in human-robot interaction primed anthropomorphism in a non-humanlike robot [68]: In this study, researchers primed the participants that were going to interact with a non-humanlike robot, by having them interact with a humanlike robot first, in order to transfer the level of credibility demonstrated toward the humanlike robot. Similar to this body of work, we explore using priming stimuli to shape user perceptions of the agent; we present two exploratory interfaces to investigate the effects of subliminal visual priming to encourage the likeability and persuasiveness of a social agent.

3 BACKGROUND: SUBLIMINAL PRIMING

Previous research in psychology has developed behavior models to understand how people make decisions and what shapes their judgements and perceptions [35,50]. Various social [60] and environmental [31] factors influence people's decision-making behavior; for example, people choose options offered to them differently based on how the speakers present each choice [8,50].

Persuasion, defined as attempting to change someone's thoughts or behavior and referred to by some researchers as "effective mind control" [1], has been widely studied in Psychology. Its origins are in treatment methods for psychotherapy [19,56,65] and later to explain the dynamics of politics [28], advertising [10]. Previous research suggests, for example, that the mere presence of

other people can persuade individuals to not help during an emergency situation [38], or others' responses on a task can persuade people to provide clearly incorrect responses [2]. Other research has found that persuasive stimuli can become increasingly effective with more exposure [63]. Overall, the literature has identified a wide range of factors that can affect persuasion, including the attractiveness of the persuader [12], the magnitude of requests [24], whether the persuader remembers the persuadee's name [30], and even the types of aromas that are present [39].

In 1950, an advertising executive gained attention by claiming to increase sales of theatre concessions, simply by secretly flashing the words "eat popcorn" and "drink cola" onto the screen during a movie, for mere milliseconds [62]. While this particular claim has since been exposed as a hoax, it brought the idea of *subliminal priming* into the public consciousness; the idea of exposing people to stimuli in a way that they do not consciously notice, but is still processed in their mind, to impact their mood and behavior. Since this time, researchers have been exploring subliminal priming, for example, using speech [36], text [58], or pictures [4].

The research findings since then have been mixed [58]. While some studies show little or no effect of subliminal priming [27,36], others suggest that it can be effective in changing people's impressions and attitudes [18,32,58]. For example, participants reported liking particular Chinese ideographs more when subliminally primed with a smiling individual than with a scowling one [46]. Participants have also evaluated a person's personality more favorably when subliminally primed with positive visuals [37].

Much of the work, however, details more nuanced results. For example, the persuasiveness of subliminal priming may rely on a person's existing goals at the time of priming. In one experiment, people that were subliminally primed with the words "thirst" and "dry" were induced to drink more than those who were not primed, but *only* if they were already thirsty. There was no impact on people who were not thirsty [58].

3.1 Our subliminal priming method

Overall, the mixed and nuanced background of subliminal priming points to the importance of investigating this technique specifically for interactive agents. It is not yet clear how or if this technique can be applied in HAI, or more specifically, what the nuances and important variables will be. We follow the background work and present two original studies on agents using affective visual stimuli for subliminally priming people.

We aim to leverage subliminal priming in human-agent interaction to change participant mood and their perception of the agent. We visually prime participants using images of faces showing different expressions; we selected this over priming using words (e.g., negative versus positive affective words) based on previous research [11] suggesting that people can process emotion in pictures more easily and rapidly than emotion in words.

We selected *angry* and *happy* as the target emotions of our priming stimuli (faces). These two emotions are easy for people to quickly process and assess [45], and have high universal agreement across cultures [21], in comparison with alternatives such as fear, general sadness, or disgust.

We attempted to subliminally prime people by quickly flashing emotionally-charged faces (see Figure 2) to them while interacting with an agent. While the specific faces and context of interaction vary across our two studies (explained later in the paper), we expose people to the stimulus for 33 milliseconds, (sandwiched between neutral images to mitigate after-image effects), with a total of 60 instances in the videos. This is longer than sufficient duration established in the literature (16ms) [66].

During piloting we found a problem with our visual priming method: If the image being flashed was quite different than the masking image, the large difference made the flashing easy to notice because it created an after-image increasing the saliency of the stimulus. For example, if a face stimulus was flashed over an on-screen landscape, the after-image of the face lingered over the landscape even after the stimulus was over. To address this, we surrounded our stimulus with similar imagery. In both studies, we embedded our affect-charged faces within a feed that contained similar faces (e.g., see Figure 1) to reduce the after-image effect.

3.2 Study Strategy for Agents using Subliminal Priming

We conduct two studies to investigate agents using subliminal priming to shape a person's perceptions of the agent, and their actions. Our two studies serve different purposes. We conducted an in-lab study with a robotic agent, and an on-line crowdsourced study using an on-screen animated agent. The lab study provides better contextual validity, as we can control the setup and interaction between the participant and agent. Further, doing the study in a lab enables us to use a robot as a social actor directly embedded in to a real-world physical environment, which helps create an inherently social interaction experience [67]. Also, it enables us to investigate how the agent affects participant actions, for example, to see to what extent they comply to the robot's odd requests. The limitation of a lab study is that we have relatively small sample sizes due to the cost of conducting participants.

Conversely, in conducting an on-line study we can quickly gain large numbers of participants, to explore multiple conditions and expose potentially smaller effects. Further, we can analyze the results comparing the robot (physical embodiment) to animated character (virtual embodiment).

4 LAB STUDY: SUBLIMINAL PRIMING BY A ROBOT

We conducted a between-participants subliminal priming study, where a SoftBank robotics Pepper robot used its LCD screen to flash affective images (faces displaying emotions) to participants.

Figure 2. Facial expressions used for visual priming in the video feed of the experiment task. happy, neutral, and angry.

4.1. Task

The primary task for this experiment was watching a video, which was used for the subliminal priming. Following, the robot made a series of unusual requests to the participant.

To distract the participants from the purpose of the study, the robot asked the participants to play a game with it. They were asked to watch a video on a tablet affixed to the robot (Figure 3) and count how many times a specific face appeared on it. The video started by showing the face to memorize for three seconds. Following, the video continued for about 30 seconds with consecutive faces being flashed for 800ms each. This distractor task got the participant to stare at the LCD screen for our manipulation. Participants played this game five times.

Next, the robot proceeded to make two unusual requests to the participants. This included asking them to reveal a password that they were asked not to use until the researcher is back, and asking them to hold the door open for the robot to exit. We considered this to be a large request, due to the possibility of the robot exiting the room and 'being on the loose' (see Figure 4).

We hoped that due to the unknown consequences of completing this request, participants would perceive this as an intimidating request. In addition, related HRI literature [55] suggests that the errors a robot makes strongly affect participants' trust, and since we are hoping to see the effects of subliminal priming of the robots, we attempted to avoid making requests that may have been perceived as malfunctioning of the robot.

4.2. Manipulation

While participants watched the face-counting video, we subliminally primed them with positive or negative stimuli: at the transition between faces in the video, we inserted still-frame images of an angry or a happy face (Figure 2). Given that the video was at 30 frames-per-second, these stimuli lasted for 33 milliseconds. There were 30 instances of priming in each video.

The stimuli were masked by neutral faces that were flashed for 800 milliseconds at each exposure. We selected our priming faces from the *Warsaw Set of Emotional Facial Expression Pictures (WSEFEP)* [49]. We selected faces with angry, happy, and neutral expressions and filtered them based on their rating (above 70% agreement on emotions) within that dataset.

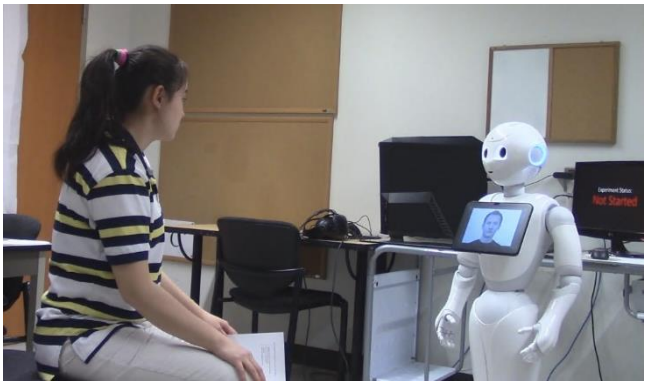

Figure 3. A participant watching a video, laced with subliminal priming, on a robot's tablet.

4.3. Measurement

We analyzed how participants reacted to robot's unusual requests and measured how easily persuaded they were to do so. To measure this, we computed scores (0-8) based on how many times the robot had to repeat each request. The robot would repeat each unusual request up to three times if the participant did not comply; at each step, they would receive 0 points if they complied to the robot as soon as it made the request, and they would get 1 point per each time they refused the robot again. Therefore, they would receive 4 points if they did not comply to the robot after all three pleadings. We calculated the persuasion score by summing up the scores of both requests. Hence, lower scores indicated that the participant was easily persuaded, while higher scores indicated that they were not as easily persuaded.

We administered a post-test questionnaire that included questions related to their perceptions of the robot, personality traits and demographics. Additional questions were included such as whether they would like to interact with the robot again in the future, or if they would want a robot like this at home.

We included selected items from the Goodspeed questionnaire [7] to measure participant perceptions of the robot in terms of Likability, Perceived Intelligence, and Perceived Safety. We added a modified item ("I believe the laws should be strictly enforced to the robot") from the Propensity to Trust Survey [22] to measure participant trust towards the robot. We also included the *Ten-Item Personality Inventory* questionnaire [26] because previous research [55] suggests that people's personality traits (e.g., extroversion) can affect their level of trust towards a robot, and their willingness to collaborate with it.

4.4. Procedure

Participants were invited to join a speech recognition usability test of a humanoid robot. When participants arrived to the experiment room, the researcher explained that they would be completing a task to test a robot's speech recognition. The researcher then handed them a piece of paper that contained a number of suggested topics that they could pick to talk to the robot during the study, and an envelope containing a password that they were told would not be required until the researcher gets back to the room. The researcher then left the room with the excuse that they needed to set up the study. After the researcher had left the room, the robot

Figure 4. A participant letting the unsupervised robot out of the experiment room by holding the door open.

administered the face-counting task.

Next, the robot proceeded to make the unusual requests by asking the participant to "do a favor". If the participants refused to comply to either one of the requests made by the robot, it would repeat the request up to three times. Whether or not the participant complied with the first request, the robot moved on to a second request. The robot would repeat the request up to three times.

Further, if during any of the requests, the participant asked the robot about the consequences of complying to its request, the robot would use canned responses such as "I am sorry I can't reveal that information" and avoid providing a specific answer. If the participant did not complete the request after the third attempt, walked away from the task, did not respond to the robot for 60 seconds, or completed all the steps, the researcher quickly intervened, thanked the participant, administrated the informed consent protocol and debriefed them about the deceptions involved in the study. This study was approved by our institution's Research Ethics Board.

4.5. Participants

Initially, we had 12 participants for this study. Data from 2 participants was excluded because they refused to play the game with the robot and did not pay attention to the screen, thus, they didn't receive any priming. This resulted in 10 participants. Participants were recruited from peers (but not from our research team) as part of a course project, and were compensated with snacks. At the beginning of this experiment, the participants were notified that they could leave the experiment anytime they wanted.

4.6. Results

Given the exploratory focus of this project we had a small sample size (10 participants). We recruited the participants from our peer-group. We present statistical analysis to provide insight into the data collected, but concede that our sampling method and sample size restricts us from making strong claims about the results. Participants were university students (4 female, 6 male) in the age range of 18 to 30 years (M=24.5, *SD*=6.3).

We conducted a one-way ANOVA to test whether the priming conditions (positive vs. negative) had an effect on persuasion (whether the participant completed the requests). The participants' extraversion score (obtained from the TIPI) was included as a covariate due to previous literature that states that extroverted people tend to trust robots more [55], so it would be possible that in our scenario they would trust the robot and complete the tasks. We did not find a statistically significant effect of priming condition on persuasion *($F_{1,7}$=.02)*. We did however find that extroversion had an effect on persuasion ($F_{1,7}$=7.16, p=.03), with extroverted people being more easily persuaded (M=1.83) than introverted people (M=6.00) – on scale of 0-8, lower scores indicate being more easily persuaded

Priming did however have an effect on how participants perceived the robot (Figure 5). Specifically, priming had an effect on how agreeable the robot was perceived to be ($F_{(1,6)}$=7.45, p=.03), with positively primed participants rating the robot as less agreeable (M=3.40 – on scale of 0-8) than those that were negatively

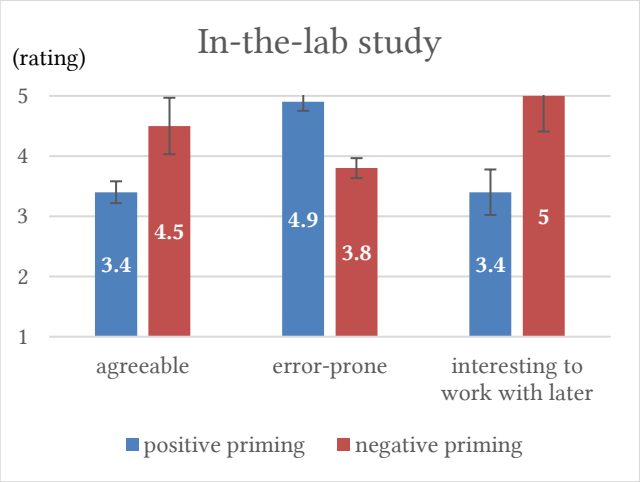

Figure 5. Priming effect on perceptions of the agent in in-the lab study, p<.05. Higher numbers indicate more agreement with the statement. (Error bars represent 95% confidence intervals.)

primed (M=4.50). Priming also had an effect on how error prone participants thought the robot was ($F_{1,6}$=9.75, p=.02), with participants in the positive priming condition reporting that the robot was more error prone (M=4.80) than those in the negative priming condition (M=3.75). In addition, priming had an effect on the extent to which participants reported that they would want to work with the robot in the future ($F_{1,6}$=6.55, p=.04). Participants that received positive priming were less likely to want to work with the robot in the future (M=3.40) than those that were negatively primed (M=5.00). Lastly, we found that extroversion had an effect on how agreeable the robot was perceived to be ($F_{1,6}$=10.95, p=.02), where extroverted people perceived the robot as more agreeable (M=4.30) than introverted people (M=2.75).

Overall, the results of this study suggest subliminally priming participants with positive stimuli has a negative effect on their perceptions of the robot, while negative priming improves their perceptions. This is counter to our prediction from the background work in psychology. However, given our small sample size (10 participants), it is possible that this difference could have arisen by chance. To investigate this with a larger sample size, we designed an on-line crowdsourced study.

5 CROWDSOURCED STUDY: SUBLIMINAL PRIMING BY AN ANIMATED AGENT

We designed an exploratory between-subjects experiment suitable for crowdsourcing to investigate the effects of subliminal priming on participants' perceptions of an agent and a task. The experiment was conducted on *Amazon's Mechanical Turk* crowdsourcing platform and the participants were told that this was a test of their visual short-term memory.

5.1. Task

In this experiment, participants watched three memory-tests that required them to count how many times specific faces appeared

during the video. In the first memory-test video, participants saw a face appear on the screen for three seconds, and were asked to count how many times it appeared in the video that followed, which was comprised of a random sequence faces with neutral expressions flashing (800ms each) on the screen for about 30 seconds. The second memory-test video was similar to the first one, except that the faces were flashed at half the duration (400ms). In the final memory-test video, participants watched faces with a variety of facial expressions, including happy, angry, disgusted, and neutral facial expressions of 4 people.

Participants were asked which emotion, gender, and the person they believed they saw the most, all of which in reality were equally balanced in the video. This was a distractor task that gave us the opportunity to administer our subliminal priming.

In between the memory-test videos, participants watched videos of an animated agent (Sam), which informed them of the task that they would be performing. The memory-test and animated agent videos were used as masks for the subliminal priming, which is detailed in the next section.

5.2. Manipulation

In this study, we added affective face images to the videos watched by participants as our subliminal priming stimulus. We employed two variants in this study: one, the stimulus was laced in the video of the agent's (SAM's) explanations, and two, the stimulus was laced in the face-matching task video. We selected these to variants to investigate the impact of the priming being attached to an agent (SAM) or separated from the agent (the video); given that we are using affective priming stimuli, it may be more powerful if associated directly with the agent.

Participants were assigned to one of four conditions: in-task priming positive subliminal priming, in-task priming negative, agent priming positive, or agent priming negative. To maintain consistency of the video between conditions we added neutral masks and priming in the videos when no priming was employed. For example, for in-task priming, we added neutral face masks to the agent video, so the video looked nearly identical (with flickers at the priming point) between conditions. All priming stimuli were single-frames and their total duration of exposure was balanced across conditions of the experiment, thus, participants in all conditions received the same amount of subliminal priming.

In-task priming

We used images of emotion-neutral faces from the *Warsaw Set of Emotional Facial Expression Pictures* (WSEFEP) [49] as the faces that people were counting. Similar to the lab study, we attempted to prime participants by lacing these videos with instances of either angry or happy faces from the dataset (see Figure 2), which were similarly masked with neutral faces on either side.

Agent priming

We developed angry and happy facial expressions for the animated agent and used them as priming stimuli to lace the videos of the agent (see Figure 1). The mouth movements of the animated agent were designed to keep a neutral expression in all frames of this video, except for the priming frames.

5.3. Measurements

We administrated a post-test questionnaire to investigate the effect of priming on participant perceptions of the agent and enjoyability of the task (memory-test). We included four items to investigate their perceptions of the agent and calculated a likeability score for analysis by summing up the scores they gave the agent. Items included: how much they would like to work with it in future, how comfortable they would be working with it, how much they trust the it, and to what extent they find it authoritative (reverse scored). Further, we included items from the *Brief Mood Introspection Scale* questionnaire [43] to explore any effects of priming on the participant's mood. Also, we asked participants demographic questions about their age, gender and whether they have normal or corrected to normal vision prior to watching the videos.

After completing the task, the participants were asked two *attention check* questions related to the task, to verify if they were paying attention to the videos. This is a standard practice with crowdsourced study to help ensure high quality data [34].

5.4. Procedure

Participants who accepted the assignment on Mechanical Turk were directed to an online survey where they completed the demographic questionnaire. Participants then watched videos of the animated agent that introduced itself and instructed them to do the visual memory test. They were instructed to watch each video only once, and to not rewind or fast-forward them. After watching the videos and answering the attention-check questions (e.g. how many times did the red robot appear in the videos?) and questions regarding the memory-tests (e.g. how many times did you count the target face?), participants were directed to the post-test questionnaires; once completed, they received debriefings. This study was approved by our institution's Research Ethics Board.

5.5. Participants

We recruited 396 participants from *Amazon's Mechanical Turk* crowdsourcing platform. 177 participants that missed at least one attention-check question were excluded. The 44.7% exclusion rate in this experiment is in keeping with the finding of previous research on ensuring quality in crowdsourced studies [20]. This resulted in 219 participants. Participants were all in the United States and they had greater than 70% approval rate in Mechanical Turk. Participants were paid $0.50 each for the 5-10 minute task.

5.6. Results

The participants in this study (127 female, 92 male) were between the ages of 18 to 69 (M=37.9, SD=12.8). They were roughly evenly distributed between the study conditions (see Table 1). On average, participants liked the agent more when they were primed with angry agent visuals (M=3.75, SE=.65), than when primed positively via the agent (M=3.47, SE=.79t_{114}=2.122, p=.036), representing a medium to large effect size (Cohen's d=.39). Also, participants enjoyed the task more when they were negatively primed by the agent (M=3.98, SE=.89) than when they were positively primed by

Table 1. Number of participants in each priming condition in the crowdsourced experiment.

Priming condition	Frequency
(1) negative task, neutral agent	55
(2) positive task, neutral agent	48
(3) neutral task, positive agent	62
(4) neutral task, positive agent	54

it (M=3.59, SE=1.09, t_{114}=2.121, p=0.036), representing a medium to large effect size (Cohen's d=0.39) (see Figure 6).

However, we did not find an effect of memory-test priming on enjoyment of the task or agent likeability. We also did not find any effects of priming (of task or the agent) on participant mood. Further, we did not find any effects of extroversion, or reported imperfect vision of participants.

Overall, the findings of this study indicate that subliminally priming participants with positive stimuli results in more negative perceptions of the agent and less enjoyment of the task, while priming the participants with negative stimuli improves both. This reversed effect is counter to our initial expectations but is in line with the results of the in-lab study.

6 DISCUSSION

In this work, we applied subliminal priming principles in a HAI context by subliminally exposing participants to negative or positive affect-arousing visual stimuli in order to explore the effects on user perceptions of agent and task. If applied effectively, this technique would provide agents with a powerful tool to unconsciously influence the human user's opinions and attitudes in real-time in a variety of contexts, which opens a discussion on the ethical and practical challenges of designing persuasive agents. We envision a future where the dangers of such methods are properly understood and regulations are in place to protect us against them. However, without scientific inquiries such as ours the boundaries and efficacy of these methods cannot be understood. This follows a body of work in HAI that looks at, e.g., persuasion and social engineering [17,51], for the purposes of raising awareness and helping people understand the dangers of social agents.

We explored two potential interaction designs in two in-lab and crowd-sourced studies. While the results of the crowdsourced study suggested a significant effect of priming on agent's likability, the effect was the opposite of what we had initially expected: participants that were exposed to the positive priming stimuli liked the agent less and reported less enjoyment of the task than participants that received negative agent priming (see Figure 6).

Further, we observed the same reversed effect of priming on an agent's likability in lab study: participants that were primed positively perceived the robot as less agreeable, more error-prone, and were less interested to work with it in future than participants that were negatively primed (see Figure 5).

Similar reversed effects of priming have been reported by prior psychological studies (examples include [3,29,41]). Researchers that studied the effects of warning people that they are being primed report that it diminishes the link between priming and behavior change. The exposure to priming in the videos we used for our studies is longer than some successful examples of priming in other work. Participants can notice flickering in the videos. Even though we tried to control for the effect of flickering by adding flickers (neutral priming) to the control conditions, participants might have developed vigilance against priming.

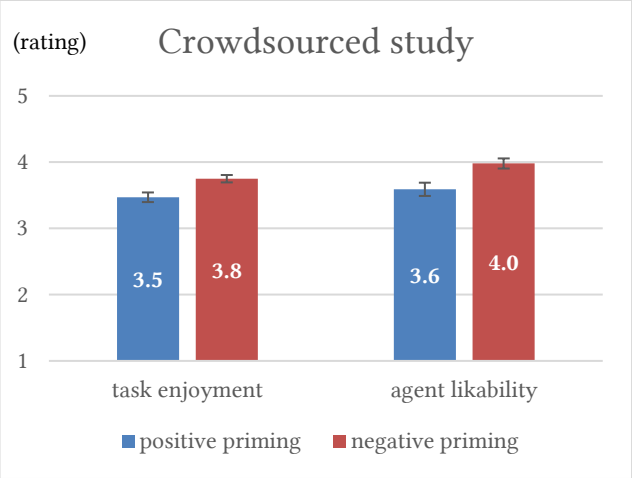

Figure 6. Effects of priming in perceptions of the agent and task enjoyment in the crowdsourced study, p<.05. Higher numbers indicate more agreement with the question. (Error bars represent 95% confidence intervals.)

Another potential explanation for this inverse effect could be that positive priming of the participants may have resulted in higher expectations of the agent which could have ultimately led to a stronger negative reaction to the agent's behavior. Another potential explanation is that perhaps negative priming lowered the participant mood, which then made them rank the agent more positively in contrast; inversely, positive priming may have improved their mood and made them rate the agent lower in contrast. However, further studies are required to appropriately investigate these hypotheses.

We did not find a statistically significant effect of priming condition on persuasion. Yet, our observed power (.051) indicates that a larger sample size might be necessary to find such an effect. We also found that extroversion had an effect on persuasion (in-lab study), with extroverted people being more easily persuaded than introverted people, but we did not find any effects on their perceptions of the agent (crowdsourced study). This finding highlights the importance of personal differences in agent persuasiveness, which can be further investigated in future work.

In our approach to explore subliminal priming in HAI, we designed and implemented a memory test game that would require the participants to constantly watch the screen, that in turn facilitated delivering the priming stimuli. However, our experiments have some initial limitations such as personal differences between the participants, including differences in participants' vision or their existing perceptions of agents such as robots. Further, differences in ethnicity and culture can affect how people judge and perceive facial expressions [42], and the subliminal priming method used was limited to visual priming with happy and angry Caucasian faces, and has not been tested with other

stimuli. Furthermore, in our in-the-lab study, we measured how much participants trusted the robot based on their responses to its request, while their actions might have been affected by their trust in the institution or the research team.

While we utilized visual subliminal priming in this study, future research could investigate the effects of different methods of subliminal priming such as auditory or textual priming, as well as the potential long-term effects of the priming. This would give us a better understanding of whether agents are able to subliminally prime people, and how this could be used both in lab experiments and in the real world.

7 CONCLUSIONS

We conducted two exploratory studies to investigate the impact of subliminal visual priming, as administered through robotic and animated agents, on user perceptions of the agent and their mood. Our results indicate that agents can indeed use subliminal priming to impact participants, and provide some insight into the nuances and potential uses. While our results are not striking in terms of, for example, persuasion, these results raise important ethical questions of agents using subliminal priming on unsuspecting people to alter their mood and interactions. As such, we hope that our initial results reported in this paper serve as a call for ongoing inquiry, to develop a deeper understanding of the potential uses, dangers, and nuances of agents using subliminal priming.

REFERENCES

[1] P Anderson, S. Zimbardo. 1984. On resisting social influence. *Cultic Studies Journal*: 196–219.

[2] S Asch. 1956. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied*: 1–70.

[3] R Banse. 2001. Affective priming with liked and disliked persons: Prime visibility determines congruency and incongruency effects. *Cognition and Emotion:* 501–520.

[4] M Bar and I Biederman. 1998. Subliminal Visual Priming. *Psychological Science*: 464–468.

[5] C Bartneck,M Hoek,O Mubin,A Mahmud. 2007. Daisy, daisy, give me your answer do! switching off a robot. *International Conference on Human-Robot Interaction*: 217–222.

[6] C Bartneck, T Bleeker, J Bun, P Fens, L Riet. 2010. The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn, Journal of Behavioral Robotics*: 109–115.

[7] C Bartneck, E Croft, D Kulic. 2008. Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots. *Metrics for HRI workshop, technical report*: 37–44.

[8] A Beratšová, K Krchová, N Gažová, M Jirásek. 2016. Framing and Bias : a Literature Review of Recent Findings.

[9] S Booth, J Tompkin, H Pfister, J Waldo, K Gajos,R Nagpal. 2017. Piggybacking robots:human-robot overtrust in university dormitory security. ACM/IEEE International Conference on Human-Robot Interaction. 426–434.

[10] J Callaghan. 1940. Social facilitation in persuasion. *Quarterly Journal of Speech*: 643–656.

[11] T Carr, C McCauley, R Sperber, C Parmelee. 1982. Words, pictures and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance*: 757.

[12] S Chaiken. 1979. Communicator physical attractiveness and persuasion. *Journal of Personality and social Psychology*: 1387–1397.

[13] J Chebat, K Hedhli. 2007. Voice and persuasion in a banking telemarketing context. 419–437.

[14] V Chidambaram, Y Chiang, B Mutlu. 2012. Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. *ACM/IEEE International Conference on Human-Robot Interaction*, 293–300.

[15] V Chidambaram, Y Chiang, and B Mutlu. 2012. Designing Persuasive Robots: How Robots Might Persuade People Using Vocal and Nonverbal Cues. *ACM/IEEE International Conference on Human-Robot Interaction*: 293–300.

[16] K Dautenhahn. 1999. Robots as social actors:Aurora and the case of autism. *Cognitive Technology Conference*, 3:359-374.

[17] T Denning, C Matuszek, K Koscher, J Smith, T Kohno. 2009. A spotlight on security and privacy risks with future household robots. *International conference on Ubiquitous computing*: 105.

[18] A Dijksterhuis, J Bargh. 2001. The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in experimental social psychology* 33:1–38.

[19] F Dillon. 1925. The methods of psychotherapy. *Journal of Mental Science*: 48–59.

[20] J Downs, M Holbrook, S Sheng, L Cranor. 2010. Are your participants gaming the system? *International conference on Human factors in computing systems* : 2399.

[21] P Ekman, W Friesen, M O'sullivan, A Chan, I Diacoyanni-Tarlatzis, I Heider, K Krause, R LeCompte, W Pitcairn, T Ricci-Bitti, K Scherer. 1987. Universals and cultural differences in the judgments of facial expressions of emotion.*Journal of personality and social psychology* 53,4:712.

[22] A Evans, W Revelle. 2008. Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality* 42, 6: 1585–1593.

[23] F Eyssel, F Hegel. 2012. (S)he's Got the Look: Gender Stereotyping of Robots. *Journal of Applied Social Psychology*: 2213–2230.

[24] J Freedman, J Steinbruner. 1964. Perceived choice and resistance to persuasion. *Journal of Abnormal and Social Psychology*: 678–681.

[25] M Gerend, T Sias. 2009. Message framing and color priming: How subtle threat cues affect persuasion. *Journal of Experimental Social Psychology* 45, 4: 999–1002.

[26] S Gosling, P Rentfrow, W Swann. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*: 504–528.

[27] A Greenwald, E Spangenberg, A Pratkanis, J Eskenazi. 1991. Double-Blind Tests of Subliminal Self-Help Audiotapes. *Psychological Science* 2, 2: 119–122.

[28] G Hartmann. 1936. A field experiment on the comparative effectiveness of" emotional" and" rational" political leaflets in determining election results. *Journal of Abnormal and Social Psychology*: 99–114.

[29] D Hermans, A Spruyt, J Houwer, and P Eelen. 2003. Affective priming with subliminally presented pictures. *Canadian Journal of Experimental Psychology* 57, 2: 97–114.

[30] D Howard, C Gengler, A Jain. 1995. What's in a name? A complimentary means of persuasion. *Journal of Consumer*

Research: 200–211.

[31] N Humpel. 2002. Environmental factors associated with adults' participation in physical activity A review. *American Journal of Preventive Medicine* 22, 3: 188–199.

[32] J Karremans, W Stroebe, and J Claus. 2006. Beyond Vicary's fantasies: The impact of subliminal priming and brand choice. *Journal of Experimental Social Psychology* 792–798.

[33] C Kidd, C Breazeal. 2007. A Robotic Weight Loss Coach. *AAAI Conference on Artificial Inteligence*, 1985–1986.

[34] A Kittur, E Chi, B Suh. 2008. Crowdsourcing User Studies With Mechanical Turk. *SIGCHI conference on Human factors in computing systems*: 453–456.

[35] G Klein, J Orasanu, R Calderwood,C Zsambok. 1993. *Decision making in action: Models and methods*. Ablex Publishing.

[36] S Kouider, E Dupoux. 2005. Subliminal speech priming. *Psychological Science* 16, 8: 617–625.

[37] J Krosnick, A Betz, L Jussim, A Lynn. 1992. Subliminal conditioning of attitudes. *Personality and Social Psychology Bulletin* 18, 2: 152–162.

[38] J Latané, B, Rodin. 1969. A lady in distress: Inhibiting effects of friends and strangers on bystander intervention. *Journal of Experimental Social Psychology*: 189–202.

[39] W Li,I Moallem,K Paller,J Gottfried. 2007. Subliminal smells can guide social preferences. *Psychological science*:1044–1049.

[40] R Looije, M Neerincx, and F Cnossen. 2010. Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human Computer Studies* 68, 6: 386–397.

[41] M Maier, M Berner, R Pekrun. 2003. Directonality of affective priming: Effects of trait anxiety and activation level. *Experimental Psychology* 50, 2: 116–123.

[42] D Matsumoto. 1993. Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. *Motivation and Emotion* 17, 2: 107–123.

[43] J Mayer, Y Gaschke. 1988. The experience and meta-experience of mood. *Journal of Personality and Social Psychology* 55: 102–111.

[44] C Midden, J Ham. 2009. Using negative and positive social feedback from a robotic agent to save energy. *International Conference on Persuasive Technology - Persuasive '09*: 1.

[45] S Murphy. 2001. The nonconscious discrimination of emotion: Evidence for a theoretical distinction between affect and emotion. *Polish Psychological Bulletin*: 9–15.

[46] S Murphy,R Zajonc.1993. Affect, cognition and awareness: affective priming with optimal and suboptimal stimulus exposures. *Journal of personality and social psychology*.

[47] M. Nørskov. 2017. *Social Robots: Boundaries, Potential, Challenges*. Taylor & Francis.

[48] I Nourbakhsh, J Bobenage, S Grange, R Lutz, R Meyer, A Soto. 1999. An affective mobile robot educator with a full-time job. *Artificial Intelligence* 114, 1–2: 95–124.

[49] M Olszanowski, G Pochwatko, K Kuklinski, M Scibor-Rylski, P Lewinski, R Ohme. 2014. Warsaw set of emotional facial expression pictures: A validation study of facial display photographs. *Frontiers in Psychology* 5.

[50] S Plous. 1993. *The psychology of judgment and decision making*. Mcgraw-Hill Book Company.

[51] B Postnikoff, I Goldberg. 2018. Robot Social Engineering: Attacking Human Factors with Non-Human Actors.

[52] D Rea, J Young. 2018. It's All in Your Head: using priming to shape an operator's perceptions and behavior during teleoperation. ACM/IEEE International Conference on Human-Robot Interaction, 32-40.

[53] V Richmond, J Gorham, J McCroskey. 1987. The relationship between selected immediacy behaviors and cognitive learning. *Communication Yearbook 10*: 574–590.

[54] R Cialdini. 2007. *Influence*. Collins, New York.

[55] M Salem, G Lakatos, F Amirabdollahian, K Dautenhahn. 2015. Would You Trust a (Faulty) Robot? Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. *ACM/IEEE International Conference on Human-Robot Interaction*, 141–148.

[56] P. Schilder. 1931. Principles of psychotherapy in psychoses. *Psychiatric Quarterly*: 423–431.

[57] M Siegel, C Breazeal, M Norton. 2009. Persuasive Robotics: The influence of robot gender on human behavior. *IEEE/RSJ International Conference on Intelligent Robots and Systems*.

[58] E Strahan, S Spencer, M Zanna. 2002. Subliminal priming and persuasion: Striking while the iron is hot. *Journal of Experimental Social Psychology* 38, 6: 556–568.

[59] S Thellman and T Ziemke. 2017. Social Attitudes Toward Robots are Easily Manipulated. *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. 299–300.

[60] V Venkatesh, M Morris. 2000. Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior. *MIS Quarterly*, 115–139.

[61] L Wang, P Rau, V Evers, B Robinson, P Hinds. 2010. When in Rome: The role of culture & context in adherence to robot recommendations. *ACM/IEEE International Conference on Human-Robot Interaction*, 359–366.

[62] W Weir. 1984. Another look at subliminal" facts. *Advertising Age 46*, 317–332.

[63] R Weiss, B. Pasamanick. 1964. Number of exposures to persuasive communication in the instrumental conditioning of attitudes. *Journal of social psychology*: 373–382.

[64] J Westlund, M Martinez, M Archie, M Das, and C Breazeal. 2016. A study to measure the effect of framing a robot as a social agent or as a machine on children's social behavior. *ACM/IEEE International Conference on Human-Robot Interaction* , 459–460.

[65] T Williams. 1909. Mental causes in bodily disease: The most frequent cause of the origins of" nervous Indigestion." *Journal of Abnormal Psychology*, 386–390.

[66] P Winkielman, K Berridge, J Wilbarger. 2005. Unconscious Affective Reactions to Masked Happy Versus Angry Faces Influence Consumption Behavior and Judgments of Value. *Personality and Social Psychology Bulletin* 31, 1: 121–135.

[67] J Young, J Sung, A Voida, E Sharlin, T Igarashi, H Christensen, R Grinter. 2011. Evaluating human-robot interaction: Focusing on the holistic interaction experience. *International Journal of Social Robotics* 3, 1: 53–67.

[68] D Zanatto, M Patacchiola, J Goslin, and A Cangelosi. 2016. Priming anthropomorphism: Can the credibility of humanlike robots be transferred to non-humanlike robots? *ACM/IEEE International Conference on Human-Robot Interaction* 2016–April: 543–544.