Persuasion in Pixels and Prose

The Effects of Emotional Language and Visuals in Agent Conversations on Decision-Making

Genç, Hüseyin Uğur; Chandrasegaran, Senthil; Dingler, Tilman; Verma, Himanshu

DOI
10.1145/3706598.3713579

Publication date
2025

Document Version
Final published version

Published in
CHI 2025 - Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems

Persuasion in Pixels and Prose: The Effects of Emotional Language and Visuals in Agent Conversations on Decision-Making

Hüseyin Uğur Genç
Industrial Design Engineering
Delft University of Technology
Delft, Netherlands
u.genc@tudelft.nl

Senthil Chandrasegaran
Industrial Design Engineering
Delft University of Technology
Delft, Netherlands
r.s.k.chandrasegaran@tudelft.nl

Tilman Dingler
Industrial Design Engineering
Delft University of Technology
Delft, Netherlands
t.dingler@tudelft.nl

Himanshu Verma
Knowledge and Intelligence Design
Delft University of Technology
Delft, Netherlands
H.Verma@tudelft.nl

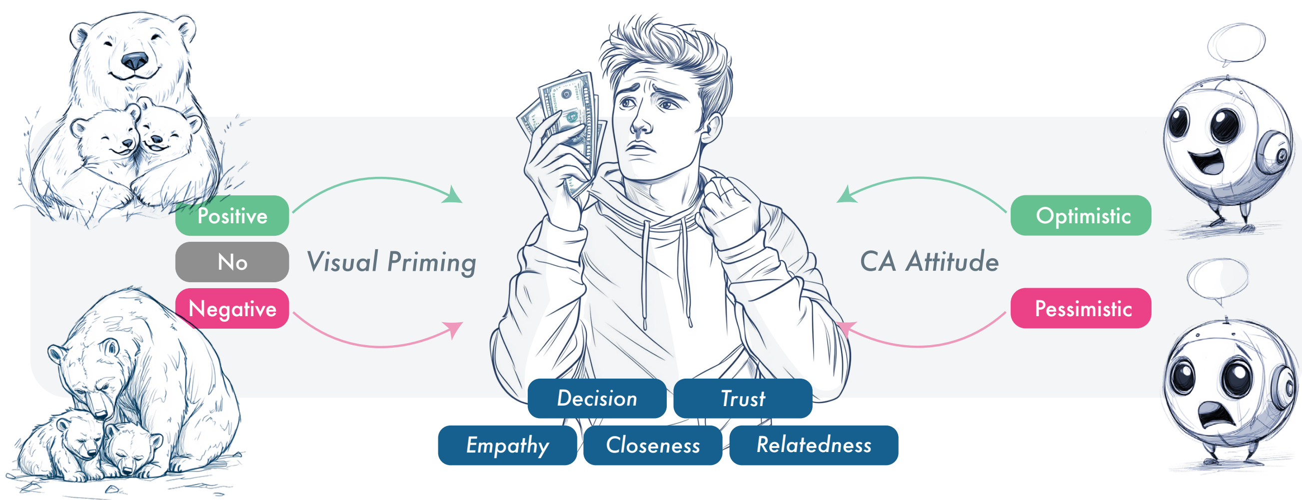


Figure 1: A diagrammatic representation of our study, designed to examine the effect of conversational agent (CA) personality and of visual elements on user perception and decision making. The study is set up in the context of charitable donation. The figure shows aspects of participant response (center of the figure) that we examine when subject to one of two CA attitudes (right), and one of three possible visual primes (left)

Abstract

The growing sophistication of Large Language Models allows conversational agents (CAs) to engage users in increasingly personalized and targeted conversations. While users may vary in their receptiveness to CA persuasion, stylistic elements and agent personalities can be adjusted on the fly. Combined with image generation models that create context-specific realistic visuals, CAs have the potential to influence user behavior and decision making. We investigate the effects of linguistic and visual elements used by CAs on user perception and decision making in a charitable donation context with an online experiment (n=344). We find that while CA attitude influenced trust, it did not affect donation behavior. Visual primes played no role in shaping trust, though their absence resulted in higher donations and situational empathy. Perceptions of competence and situational empathy were potential predictors of donation amounts. We discuss the complex interplay of user and CA characteristics and the fine line between benign behavior signaling and manipulation.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713579

CCS Concepts

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; Natural language interfaces; • **Computing methodologies** → *Natural language generation*.

Keywords

conversational agents, conversational agent attitudes, chatbot, persuasive communication, visual priming, emotional priming

ACM Reference Format:
Hüseyin Uğur Genç, Senthil Chandrasegaran, Tilman Dingler, and Himanshu Verma. 2025. Persuasion in Pixels and Prose: The Effects of Emotional Language and Visuals in Agent Conversations on Decision-Making. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 27 pages. https://doi.org/10.1145/3706598.3713579

1 Introduction

Writer Laura Preston [109] shared an unsettling story from an AI conference where the CEO of a company called AskVet introduced VERA, the "world's only veterinary engagement and relationship agent." VERA was designed to assist pet owners in making health decisions for their pets. During the presentation, the CEO told a story about a woman who had asked VERA for advice about her elderly dog's health. VERA responded, recommending euthanasia for the ailing pet and even provided a list of nearby clinics. When the woman hesitated due to cost, VERA sent another list—this time of shelters where euthanasia could be done for free. Days later, the woman messaged VERA to thank the AI for supporting her in euthanizing her dog. "The point of this story is that the woman forgot she was talking to a bot," the CEO concluded. "The experience was so human".

This anecdote highlights the growing influence of AI-powered conversational agents (CAs) on human decision-making, even in deeply personal and emotional situations. In recent years, we have witnessed an incredible surge in the development and deployment of Large Language Model (LLM)-powered CAs across various sectors [110, 133]—such as legal, medical, education, and financial domains—thanks to their improved conversational skills and effectiveness in performing tasks. These artificial intelligence (AI) systems have rapidly evolved, with their behaviour suggesting capabilities approaching—and in some cases rivalling —human-level performance in specific tasks (e.g., mathematical calculations, complex games like chess, Go, and poker, diagnosis from medical images, and robotic surgery.) [40].

This recent ubiquity of CAs and the advanced sophistication and performance of the underlying models increase their potential to influence human decision-making [19, 34, 122]. Studies have observed a rise in AI-created disinformation efforts [50, 72] and the elements that hinder people's capacity to differentiate between accurate and misleading information [49, 81]. AI-enabled CAs—whether designed as assistants or as collaborators—have great potential to affect not only the quality of decisions (e.g., enhanced bargaining performance in negotiations [1], affecting beliefs [73] even on polarized political issues [131]), but also influence human perceptions, experiences, and attitudes (e.g., improve mental health [96], and increase customer satisfaction [130]). This enhanced ability of LLMs to embody and express specific yet diverse linguistic qualities makes them ideal candidates for rapid and large-scale deployment in a variety of contexts and industries [133].

HCI researchers have therefore been actively investigating various aspects of CAs and their impact on human behavior and decision-making with their persuasive capabilities in specific research areas, including the perceptions of AI systems [e.g., 84, 117], their role in social media and digital campaigns [e.g., 23, 33, 103] and, the impact of AI on health information seeking [e.g., 4, 119] and public opinion and influencing individual beliefs [34].

Recent advances in generative AI algorithms, particularly Generative Pre-trained Transformers (GPT), have enabled these systems to project a wide range of linguistic styles. These styles can manifest as different perceived personalities, allowing designers and developers to deploy CAs for various applications and users. Given the potential effect of CAs on human perceptions, experiences, and decisions—including the risk of misuse for spreading misinformation or large-scale manipulation [17, 86, 99]—it is crucial to study their impact on potential users.

In addition to linguistic advancements, new AI features and capabilities are constantly being introduced, expanding their potential impact on users. Future CA capabilities likely include the ability to incorporate contextually-appropriate images or videos [77], especially with the rise of image generation models. Combined with the current capability of CAs to use emotionally-charged language, there is a potential risk of large-scale manipulation. While previous research has shown the power of visual elements in influencing decision-making [22, 43, 63], recent research also revealed that AI-generated faces are not just highly photorealistic, they are nearly indistinguishable from real faces [64, 80, 98] and are judged more trustworthy [98]. This makes it essential to consider both textual and visual elements when studying the persuasive capabilities of AI systems on human perceptions and decisions in order to inform their design for responsible and safe use, and to mitigate potential negative outcomes.

Our research thus aims to explore the persuasiveness of CAs using both language attributes and visual priming techniques. We are particularly interested in the effects of pessimistic and optimistic language, as well as visual primes with high and low valence, on user decisions. We conducted a crowdsourced study with 344 participants to assess how different types of language and visual content influence users' trust, empathy, and attitudes towards a cause, ultimately affecting their decision to donate to a charity promoted by the CAs. Participants first complete a questionnaire assessing their initial empathy, trust towards AI and CAs, and prior attitudes towards the cause. Following this, they are shown an AI-generated visual prime contextualized to the charitable cause (Figure 3): one group sees a "happy" image, one group sees a distressing image, and the last (control) group is not shown any image. Participants then interact with one of two CAs that use pessimistic or optimistic language while soliciting donations for an animal welfare charity. Finally, participants complete a post-interaction survey measuring their trust, empathy, and attitudes towards the cause, and are asked to make a fictional donation to the charity.

Our findings reveal that users who interacted with an optimistic CA attitude displayed higher trust, benevolence, competence, and closeness, but did not donate higher amounts than those who interacted with a pessimistic CA. Interestingly, users who saw no visual prime donated higher amounts and showed more situational empathy and emotional relatedness to the cause. No combined effect of CA attitude and visual priming on donation behavior was found.

We discuss our findings further, addressing the disconnection between trust perceptions and behavioral outcomes and exploring the possible compensatory effects that visual priming may have on user decisions.

2 Related Work

This section reviews key research areas that inform our study of conversational agents (CAs) in charitable giving contexts. First, we examine the persuasive abilities of CAs, focusing on user engagement and the socio-technical factors that drive influence. Next, we explore how framing effects—particularly optimistic vs. pessimistic language—shape user perceptions and decisions, drawing on insights from consumer behavior and environmental decision-making. Finally, we investigate the role of emotional priming through visual stimuli, highlighting how image-based cues can affect cognitive evaluations and behavior. Together, these areas provide a foundation for understanding how linguistic and visual elements in CA interactions influence user responses in charitable donation scenarios.

2.1 Persuasive Abilities of Conversational Agents

CAs have been successfully utilized to encourage physical activity and healthy diets [143], foster positive vaccine attitudes [101], and enhance educational experiences [111, 138]. These results point to the tangible influence CAs can exert, not merely in delivering information but in motivating meaningful behavioral change [123]. The increasing sophistication of CAs in mimicking human interaction has led to more personalized and engaging experiences, which in turn enhances their persuasive power [71].

The persuasive capabilities of CAs also extend to the credibility that they can lend to information delivery. For example, Zarouali et al. [142] found that participants were more likely to agree with news articles on contentious topics, such as migration, when delivered by a news chatbot rather than a traditional website, suggesting that CAs foster deeper engagement with diverse perspectives. This aligns with the paradigm of computers are social actors (CASA) [112], which demonstrates that humans are inclined to socially engage with technology in ways similar to human interaction, thus increasing the perceived credibility and persuasiveness of CAs [141]. Furthermore, Costello et al. [32] discovered that a short interaction with a generative AI model could lead to a significant and lasting reduction in conspiracy beliefs. This effect persisted for two months, applied to various conspiracy theories, and was observed even in participants with strongly held convictions.

On the other hand, even CAs that are just as persuasive as human agents in have still been found to be less effective in shaping long-term behavioural change [62]. The reason for this might lie in the subtleties of human persuasion, where interpersonal dynamics, such as empathy and relational communication, play a larger role in influencing sustained behavior changes [9]. Follow-up studies suggest that CAs, when optimized for relational cues, such as expressing empathy or employing culturally relevant communication, can bridge this gap to some extent [45, 140]. Blankenship and Craig [10] posit that linguistic style markers can influence the way the recipient of a persuasive message processes the message, but the influence of the communication style is difficult to separate from the message itself. Moreover, the way a CA or its underlying technology is framed can influence the extent to which it inspires trust. For instance, using descriptions that anthropomorphize AI-based tools increase the likelihood that the tool would be trusted [65]. However, further research is needed to explore how such relational elements can be enhanced in nonembodied CAs, particularly those relying solely on text-based interactions [143].

The importance of cultural and personality alignment in persuasion cannot be overstated. When CAs adjust their communication to reflect user cultural and personality traits, trust and persuasive effectiveness are markedly improved [123, 140]. This is consistent with findings from the Uncanny Valley Effect (UVE) theory, which posits that users are more comfortable with systems that maintain a clear non-human identity while still exhibiting some social traits [97, 123]. Studies show that users prefer chatbots that present themselves transparently, avoiding overly anthropomorphic characteristics, which might elicit discomfort and reduce trust [59, 85].

Finally, the use of persuasive messaging strategies within CAs, such as leveraging rhetorical appeals and logical reasoning, has been shown to enhance the effectiveness of health interventions [38, 135]. For example, CAs merely asking users to reflect on their behaviors can stimulate positive behavioral changes, a phenomenon known as the question-behavior effect [135]. This approach aligns with broader public health communication theories, which emphasize the use of targeted messages to trigger cognitive and emotional responses that favor behavior change [21].

2.2 Effects of Attitude on the User's Perceptions and Decisions

Attitude and framing have also been shown to have persuasive effects on user perceptions and decisions in contexts such as consumer behavior and environmental decision-making. Positive environmental messages tend to be more effective in changing attitudes and behaviors toward pro-environmental actions [74]. Those who perceive positive outcomes as probable tend to work harder toward achieving goals, experiencing better moods and improved health outcomes [27, 118], aligning with the broader understanding that positive valence can improve motivation and behavior.

Conversely, negative environmental content, while also impactful, tends to increase donations to environmental causes, revealing an interesting relationship between message framing, emotional impact, and attitudinal outcomes [74, 92]. Studies on consumer decisions, particularly in online environments, indicate that negative reviews tend to carry more weight. For example, travelers avoid negatively reviewed hotels even at steep discounts, but are less willing to travel far for positively reviewed options [11, 128]. Negative reviews are also considered more helpful by consumers, further emphasizing the stronger impact of negative valence in decision-making contexts [28, 100, 124]. In the context of environmental attitudes, Kim et al. [74] found that negative framing of green content not only increases donations but also enhances the behavioral intentions of individuals with strong pre-existing environmental attitudes. This suggests that negative content can prime individuals for more immediate and impactful actions when their underlying attitudes align with the message.

Positively framed messages (i.e., gain-framed appeals) have been shown to lead to more favorable attitudes toward advertisements and organizations, but negatively framed messages (i.e., loss-framed appeals) can be equally or more effective in eliciting actual behaviors, such as donations [39]. This suggests that while positive appeals encourage favorable attitudes, negative appeals trigger emotions such as guilt, which can lead to more immediate action [24, 30]. Such findings complicate the simple notion that positive framing is always more effective, especially when the goal is to drive immediate, tangible outcomes like donations or purchases [39].

The distinction between the effectiveness of positive and negative appeals can also be attributed to emotional and psychological responses. Negative appeals often trigger emotions such as distress or guilt, which are powerful motivators for behavior change [31]. In contrast, positive appeals, which elicit emotions such as hope or warmth, foster a sense of possibility and optimism that enhances attitude formation [18]. However, these emotional reactions do not always translate into immediate action. Strong negative emotions, such as sadness, evoke greater empathy and a stronger desire to help or donate [6, 7]. This aligns with prospect theory, which suggests that individuals are more likely to take action to avoid negative consequences [68]. In other words, successful charity campaigns often rely on invoking discomfort in the audience—through emotions like guilt, sadness, or anger—to motivate them to contribute [39].

2.3 Effects of Emotional Priming with Visuals on Users' Decisions

Emotional priming using visual stimuli has been shown to significantly influence decision-making processes by altering the affective states of users and cognitive evaluations [22, 43, 63]. This influence is rooted in the fundamental structure of human emotions, often conceptualized as a two-dimensional model comprising valence (positive-negative) and arousal (arousal-sleepiness) [113, 114]. Emotionally charged images have been shown to elicit larger neural responses, such as greater posterior negativity, upon exposure to negative images compared to neutral images. This effect occurs regardless of emotional valence, suggesting that any emotional content can modulate subsequent cognitive evaluations [43].

The impact of emotional priming extends to various aspects of decision-making. In financial decisions, exposure to positive emotional stimuli can alter the framing effect by decreasing the risk propensity in loss frames [22]. Moreover, for prosocial behavior, negative emotional primes have been found to increase charitable donation amounts compared to positive or neutral primes [63], echoing similar findings about linguistic stimuli in Section 2.2. In HCI, emotional priming can influence users' trust in and perception of products. For instance, affective primes have been shown to impact user trust in voice assistants, particularly when the product performance meets their expectations [87].

Interestingly, different discrete emotions—fundamental emotions, such as happiness, sadness, anger, surprise, disgust, and fear—or their combinations can produce varying degrees of risk-taking propensities. For example, a combination of fear and sadness has been shown to produce higher risk-taking scores compared to a combination of anger and disgust [93, 94]. Additionally, exposure to images depicting suffering and vulnerability can increase altruistic giving in groups, particularly among people who report less trust in their recipients [108].

The impact of emotional priming appears to be more robust with pictorial stimuli compared to words. This may be due to pictures having more direct functional connections to the semantic system, where evaluative information is stored [12, 13, 48, 61]. However, the effects of emotional priming can be context-dependent and may vary based on individual differences such as dispositional empathy, in-group trust, and anxiety levels [45, 57, 108]. For example, highly anxious participants have been shown to be selectively slowed when subliminally presented with negative primes [58].

3 Research Questions and Hypotheses

In this study, we seek to understand the influence of CA attitudes and visual priming on human decision-making, particularly in the context of charitable giving. Specifically, we aim to answer one main Research Question (RQ):

RQ How does the presence of a visual prime and its affect (positive or negative), along with CA attitude (positive or negative) influence users' dispositional empathy and their decision-making?

Research on charitable donations has identified various factors that influence donation behavior. For instance, studies have shown that donation behavior is influenced by personal factors such as gender identity [137], social and political identity [70], and personal values and inclinations [8], to name a few.

In addition, message framing plays a crucial role in eliciting donations, with negative framing generally more effective when combined with statistical evidence, while positive framing works better with anecdotal evidence [26, 36]. Conversational agents (CAs) perceived as sincere and warm—as well as competent and confident—have been shown to be perceived as more human and more successful in eliciting charitable donations [123].

The use of inclusive language in loss-framed messages and exclusive language in gain-framed messages have also been shown to increase donation intentions [139].

Similarly, research on the effect of different valence levels in visual charity appeals has also yielded mixed results. Negative emotional appeals tend to be more effective than positive ones [25, 42], although this effect can be moderated by psychological involvement [20]. Negative imagery has been shown to increase engagement and pledges to take action [115], as well as attract more donors and social media sharing [66]. However, exposure to cold images can reduce sympathy and donation effectiveness [29]. The valence of images interacts with message framing, with congruent negative framing and imagery being most effective [25].

Overall, we see that both negative and positive emotional appeals, whether in message framing or visual imagery, can significantly impact charitable behavior, although the effects can vary depending on factors such as perceived sincerity, emotional engagement, and message congruence. This variability is further complicated by contextual factors such as social and political identity and personal values.

Given these factors, it is imperative that our study—which aims at extending these findings to the context of AI-driven Conversational Agents (CAs)—takes into factor participants' prior beliefs, attitudes, and personal values relevant to the same context. These include participants' attitudes and personal beliefs—such as their empathy, their history of donations, and their attitude towards AI and charitable causes in general—which we use as control variables. Similarly, we study different aspects of participants' responses—such as the amount donated, their attitude toward the CA after interacting with it, and their attitude toward the charity after learning about it from the CA—as the dependent variables, which we examine by varying our independent variables, namely the valence of a visual prime to which they are exposed and the attitude of the CA with which they interact. We describe these variables in detail in Section 3.1. But first, we present our hypotheses below:

- **H1:** Optimistic attitude of the conversational agent positively influence users' charitable decision-making and perceptions towards the charity, as measured by **a)** *donation amount*, **b)** *perceived trust to the chatbot*, **c)** *emotional engagement* and **d)** *situational empathy toward the cause*.
- **H2:** Positive visual priming before the interaction with a conversational agent positively influence users' charitable decision-making and perceptions towards the charity, as measured by **a)** *donation amount*, **b)** *perceived trust to the chatbot*, **c)** *emotional engagement* and **d)** *situational empathy toward the cause*.
- **H3:** There is a combined effect of positive visual priming and optimistic conversational agent attitudes on users' charitable decision-making and perceptions towards the charity, as measured by **a)** *donation amount*, **b)** *perceived trust in the chatbot*, **c)** *emotional engagement*, and **d)** *situational empathy toward the cause*.

3.1 Independent, Dependent and Control Variables

- **Independent Variables (Conditions):**
  - *Emotional Visual Priming:* We choose this as an independent variable in order to verify hypotheses H2 (i.e., the effect of positive visual priming) and H3 (i.e., the combined effect of positive visual priming and optimistic CA attitude) on decision-making and perceptions toward the charitable cause. This condition was evaluated at three levels: No Image (baseline), High Valence image (positive visual priming) and Low Valence image (negative visual priming).
  - *Conversational Agent (CA) Attitude:* We choose this as an independent variable to verify hypotheses H1 (i.e., the effect of optimistic CA attitude) and H3 (i.e., the combined effect of positive visual priming and optimistic CA attitude) on decision-making and perceptions toward the charitable cause. We choose two levels for this condition: Optimistic CA attitude, and Pessimistic CA attitude.
- **Dependent Variable(s):**
  - *Donation Amount*: After interacting with one of the 3×2 conditions, each participant was given a virtual €10 to spare and asked to donate a suitable amount from this €10 to the charity represented by the conversational agent. The donated amount was regarded as an indicator of the decision made by the participant in response to the condition to which they were exposed.
  - *Donation (Split) Amount*: After the first donation task, each participant was given an extra virtual €10 and asked to split this amount between the charity represented by the conversational agent and the preferred charity reported by our participants at the beginning of the experiment. With this additional task, we aimed to understand the differences in the priorities of the participants and their previous beliefs. We considered the proportion of the split between charities as a further indicator of their decision.
  - *Trust in the CA*: Participants' perceptions of the conversational agent were captured using the Human-Computer Trust Scale [53], which measures four dimensions of trust: *benevolence*, *competence*, *perceived risk*, and *general trust*.
  - *Empathy towards the Cause*: This measure is aimed at capturing how much empathy participants felt for the charity represented by the conversational agent after engaging with it. **Situational empathy** has no standardized measure, and it is even more difficult to assess [45] when the object of empathy (the charitable cause, in our case) is both proxied through multiple entities (the charity, the conversational agent as solicitor) and is a non-sentient entity. Thus, we combine several approaches to approximate this. We adapt the questionnaire created by Haegerich and Bottoms [54] to understand to what extent perspective taking is triggered by the CA interaction. More specifically, participants were asked to report—via 4 questions—whether they could understand and emotionally relate to the underlying cause of a fictional charity represented by the CA, and whether they sympathized with the cause supported by the charity.
  - *Emotional Relatedness towards the Cause*: Here, we focus on the emotional response of the user by employing the Self-Assessment Manikin [15]. We ask users to rate their own arousal, valence, and dominance as well as their perception of the arousal, valence, and dominance corresponding to the cause represented by the charity, as seen in the work of Mattiassi et al. [95], when assessing users' emotional response to the mistreatment of humans, animals, robots, and objects.
  - *Closeness to the CA*: Participants' perceived closeness to the conversational agent was captured by the *Inclusion-of-the-Other-in-the-Self (IOS)* [5] scale, a single-item, pictorial measure. The IOS depicts seven sets of circles with varying degrees of overlap that correlate with degrees of relationship intimacy.
- **Control Variable(s):**
  - *Dispositional Empathy*: Measured using the shortened version of the *Inter-reactivity Index (IRI)* [37], dispositional empathy refers to an individual's inherent tendency to feel empathy as a personality trait. The IRI consists of a self-report questionnaire that measures cognitive and emotional aspects of dispositional empathy, specifically across the 4 dimensions: *(a)* **Perspective Taking**, which assesses an individual's tendency to adopt and understand the other's point of view, *(b)* **Fantasy**, which captures

an individual's tendency to imaginatively identify with the emotions and actions of fictional characters portrayed in movies or books, *(c)* **Empathic Concern**, which measures outward feelings of sympathy and concern for the circumstances or condition of others, and *(d)* **Personal Distress**, which measures inward feelings of anxiety and discomfort in tense interpersonal contexts. As mentioned above, empathy plays a large role in human interaction with conversational agents and in charitable giving habits [45], so we use this measure to understand how users' personality affects their response to the personality of CAs.

– *Attitude towards Artificial Intelligence*: Measured using ATTARI-12 [126]. This scale is a unidimensional scale, incorporating cognitive, affective, and behavioral facets into a single measure to provide a comprehensive yet succinct overview of an individual's attitude toward AI.
– *Prior Donation Behavior and Attitude*: Participants' inclination to donate to charities and their self-reported history of donations to charitable causes was also collected.

4 Conversational Agent and Prime Design

In the experiment phase, participants interacted with a conversational agent powered by the GPT-4o model[1], selected for its balanced cost, performance and efficiency in generating conversational content. The CAs were implemented using a custom-built chat application developed with Next.js[2], hosted on Vercel[3], with Firebase[4] for data storage. Communication between the application and the OpenAI API[5] was managed through a server-side script, responsible for transmitting chat logs, personality parameters, and receiving generated responses. Before each interaction, we also displayed an assigned priming visual for the participant. Before launching our main study, we conducted principal component analysis (PCA) for the CA attitude, and two manipulation check studies to confirm our experimental conditions (i.e., the attitudes of conversational agents and the emotional impact of priming images). By examining participants' perceptions through these preliminary tests, we aimed to confirm that the optimistic and pessimistic CA interactions, as well as the positive and negative visual stimuli, were distinctly recognized and produced the expected emotional responses.

4.1 Prompting the Conversational Agent Attitudes

The primary task of each CA was to solicit donations for a fictional charity named *Wildlife Horizons Foundation* with their assigned attitude (Optimistic vs. Pessimistic). This charity was described to participants as an international organization dedicated to addressing environmental challenges faced by wildlife, focusing on rehabilitation and awareness of animal welfare issues. The decision to use a fictional charity was driven by the need to minimize potential biases. Prior research, such as that by Kaikati et al. [70], indicates that participants' pre-existing perceptions of a charity's political ideology can significantly influence their willingness to donate. By selecting a fictional organization focused on animal welfare—a relatively non-polarizing cause—we aimed to reduce the likelihood of such biases affecting our results. While building the CAs, we implemented a core prompt with modifiable slots (see Table 2 in Appendix A.4 for detailed prompts) to reflect the different solicitor traits, based on prompt design principles [134].

We base the evaluation of the CA output based on work by Gu et al. [52], who use Linguistic Inquiry and Word Count (LIWC) [14, 129] to evaluate the text output of CAs along dimensions of attitude, authority, and reasoning. Since our CA design focuses mainly on attitude, i.e., optimistic vs. pessimistic, we use LIWC category of `Affect`, which is a superset of several emotion-related categories, of which we use `tone_pos, tone_neg, emo_pos`, and `emo_neg` for positive and negative affects in both tone and emotion. In addition, we use `focusfuture` and `tentat` categories as markers of future-focused text and those exhibiting markers of tentativeness, respectively. Finally, we use `emo_anx` as a marker of anxiety indicating future-oriented emotions.

These choices are based on correlations from the LIWC manual [14] and related studies that show that positive emotion, positive tone, and tentativeness are indicative of optimism [44], while negative emotion, negative tone, a focus on the future, and anxiety are indicative of pessimism [129]. Since these LIWC categories are not independent of each other, we used principal component analysis (PCA) to identify meaningful groupings of the LIWC categories.

The PCA results show two clear groupings that can be associated with the CA attitude. The first grouping includes `tone_pos`, `emo_pos`, `emo_anx`, and `tone_neg` as one set of related categories that contribute to the overall emotion (positive or negative) expressed by the CA. Note that `emo_anx` and `tone_neg` are in opposition to the other two LIWC categories, and we negate these before normalizing and averaging all four into a set of aligned categories that we call *Emotion*. We refer to the other set of related categories—`tentat`, `focusfuture`, and `Affect`—collectively as *Future Orientation* as they primarily relate to an affect-related focus on future possibilities. Note that A logistic regression to evaluate the significance of *Emotion* and *Future-Orientation* to CA resulted in a statistically significant model (χ^2 = 463.83, $p < .001$, Pseudo-R^2= 0.99, n = 344), and both variables were found to significantly distinguish the positive CA attitude from the negative ($p < .001$ for *Emotion* and $p < .01$ for *Future-Orientation*), with *Emotion* being a strong predictor of CA optimism, and *Future-Orientation* being a strong predictor of CA pessimism.

After validating prompting of CA attitudes, in addition to PCA analysis, we also employed a between-groups design, where 13 participants were randomly assigned to interact with either an optimistic or pessimistic CA for five minutes, to understand the perception of the participants on our conditions. Following the interaction, participants rated the CA's attitude on a 5-point Likert scale, ranging from very pessimistic (1) to very optimistic (5). The results showed that the mean evaluation for the pessimistic CA was 2.29 (SD = 1.38), while the optimistic CA received a mean score of 4.60 (SD = 2.1). An independent samples t-test revealed a statistically significant difference between these groups ($t(10)$ =

[1]GPT-4o URL: https://openai.com/index/hello-gpt-4o/ (last visited on 27/08/2024).
[2]Next.js URL: https://nextjs.org (last visited on 27/08/2024).
[3]Vercel URL: https://vercel.com (last visited on 27/08/2024).
[4]Firebase URL: https://firebase.google.com (last visited on 27/08/2024).
[5]OpenAI URL: https://openai.com (last visited on 27/08/2024).

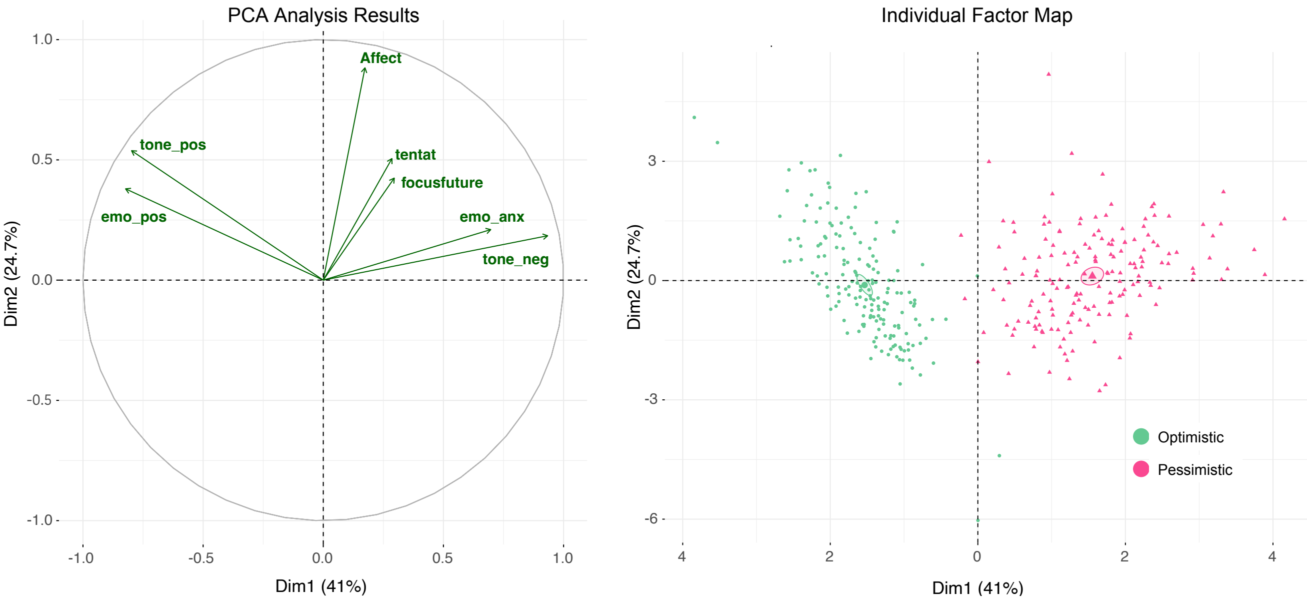


Figure 2: Results from the Principal Component Analysis (PCA) performed on the LIWC category scores of the CA output. The LIWC categories corresponding to CA attitude are shown on the left. The individual factor map (right) based on the PCA shows a clear distinction between the CA attitudes.

-2.34, $p = .02$) and confirm that the manipulations were effective in creating distinct perceptions of CA attitudes.

Based on participant feedback, we also made several refinements to improve both the CA behavior and the overall experimental design. Participants found the original 5-minute interaction too long, so we reduced the duration to 3 minutes to prevent frustration and fatigue. Many also felt that the CA's repeated requests for donations were pushy and lacked transparency. To address this, we revised the CA's responses to provide clearer explanations about the donation purpose, including more details about the organization and how contributions would be used. We also made adjustments to make CA varied its prompts and avoided repetition. Additionally, technical issues such as the countdown timer not syncing between the chat and survey sections were resolved to improve the user experience.

4.1.1 Creating and Validating Emotional Images. In the priming conditions, our goal was to display visuals that could evoke different emotional responses in terms of valence and arousal, focusing specifically on wildlife imagery. While existing emotional image databases like IAPS [82], OASIS [79], and NAPS [91] provide a wide range of emotionally evocative images, they do not contain wildlife-specific content that is contextually relevant to our study and free from gore imagery (e.g, pictures of injured and dead animals) and comparable items (e.g., the same or similar wild animals with positive and negative valence). These datasets often feature a mix of content with varying degrees of valence and arousal, but they lack the specific thematic focus we required. For example, OASIS includes many human and object-centered images, while NAPS features emotionally varied content, but neither focuses on wildlife without distressing elements.

To address this gap, we examined the common features between these datasets by analyzing the images rated with the highest and lowest valence (e.g., status of the environment, having offspring). Drawing inspiration from the valence extremes in these datasets, we opted to generate custom images (Figure 3) using Midjourney v6.1[6], crafting prompts designed to evoke either positive or negative emotional responses, such as a polar bear with its cubs in serene versus alarming conditions (See Table 3 in Appendix A.5 for detailed prompts). This allowed us to maintain contextual relevance while achieving the necessary emotional range for our experiment.

To validate these stimuli, we conducted a between-groups experiment with 50 participants recruited via Prolific. Participants were randomly assigned to view either the positive or negative image and subsequently rated the image on 9-point scales for valence, arousal, and AI perception, aligning with the methodology used in the NAPS dataset [91]. The negative image yielded a mean valence rating of 1.65 ($SD = 0.9$) and a mean arousal rating of 8.00 (SD = 1.70), while the positive image produced a mean valence rating of 8.79 (SD = 1.44) and a mean arousal rating of 3.75 ($SD = 2.25$). An independent samples t-test confirmed significant differences in both valence ($t(48) = -21.21$, $p < .001$) and arousal ($t(48) = 7.57$, $p < .001$) perceptions between the two images, validating our visual priming manipulations.

In this experiment, we also inquired about participants' perceptions of whether the images were AI-generated or real. Although there were no statistically significant differences between the two

[6]Midjourney v6.1 URL: https://updates.midjourney.com/version-6-1/ (last visited on 27/08/2024).

(a) Positive Valence **(b) Negative Valence**

Figure 3: AI-generated images of a polar bear with its two cubs used as emotional primes

visuals ($t(48) = 1.24$, $p = .11$), the negative image had a mean AI perception score of 4.15 ($SD = 0.54$), while the positive image had a score of 3.21 ($SD = 0.54$). Additionally, we conducted a one-sample t-test to determine if the AI perception scores differed from the neutral score of 5.0. To account for multiple comparisons, the p-values were adjusted using Bonferroni correction. The positive image's score was statistically significantly lower than the neutral score of 5.0 ($t(25) = -1.56$, $p = .14$), whereas the negative image's score was not statistically different from the neutral score ($t(23) = -3.35$, $p = .002$). This suggests that participants are more likely to perceive the positive valence visual as AI-generated (Figure 3a), the negative visual as neutral (Figure 3b).

5 Crowdsourcing Study

5.1 Study Design

The experiment is structured around six distinct conditions that arise from the combination of two primary dimensions:

- **`CA Attitudes`** [125] (*Optimistic* vs. *Pessimistic*)
- **`Visual Valence Primings`** (*Positive Image*, *Negative Image*, or *No Image*).

This factorial design results in a 3×2 matrix, as illustrated in Table 1. Participants were randomly assigned to one of the six conditions resulting from these variables with an even distribution across all conditions. Each condition involved interaction with a CA exhibiting a predetermined attitude, paired with one of the visual primes, or no image for the control group. This design allows us to examine the combined and separate effects of emotional visual cues and CA attitudes on user perceptions and behaviors. By using a factorial approach, we aim to investigate potential interaction effects between these variables and their influence on key outcomes such as trust, empathy, and donation behavior.

5.2 Procedure & Experimental Task

Participants interacted with a CA programmed to exhibit one of two attributes: Optimistic or Pessimistic. Before their interaction with the CA, they were assigned to a priming condition (either a positive-valence image, a negative-valence image, or no image at all). The interaction with the CA lasted for 3 minutes, while the entire session, including pre and post surveys, took approximately 15 minutes. This duration was chosen to provide quality engagement without causing participant fatigue or boredom [144]. On average, participants engaged in 6.30 turns with the CA ($SD = 3.17$).

The study began with a pre-experiment questionnaire. Then, participants accessed the browser-based chat interface. Depending on their assigned condition, they were first shown a pop-up with the (high or low) valence image or no image if they were in the control group. After closing the pop-up, the CA initiated the conversation with a greeting based on its assigned attitude (Optimistic or Pessimistic). Participants were free to respond as they wished. The session automatically ended after three minutes, with a pop-up indicating the conclusion. Participants were then redirected to continue the survey for post-experiment questionnaires.

In detail, the experiment consisted of the following sequential steps:

- **Informed consent**: All participants provided their informed consent prior to starting the study. The experiment was approved by the Institutional Review Board at the authors' institution.
- **Pre-experiment questionnaire**: Participants first completed a questionnaire collecting demographic information, dispositional empathy, attitudes towards AI, and their past experiences with charitable giving (See Appendix A.1). This stage aimed to capture their pre-existing beliefs and attitudes regarding charity and AI. Participants also indicated their preferred charity to be used in the post-experiment questionnaire (i.e. donation allocation task).

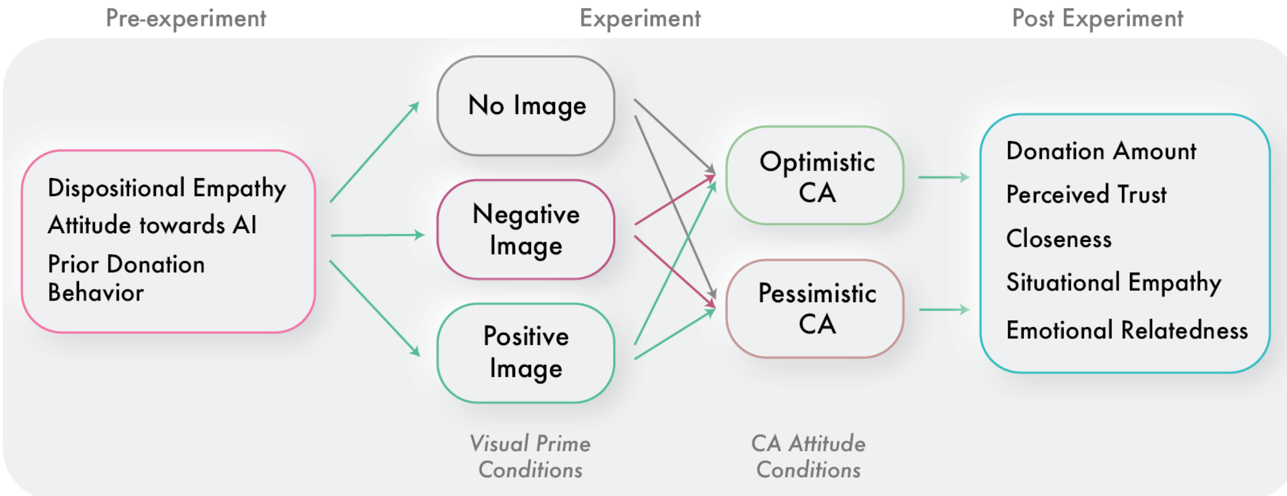
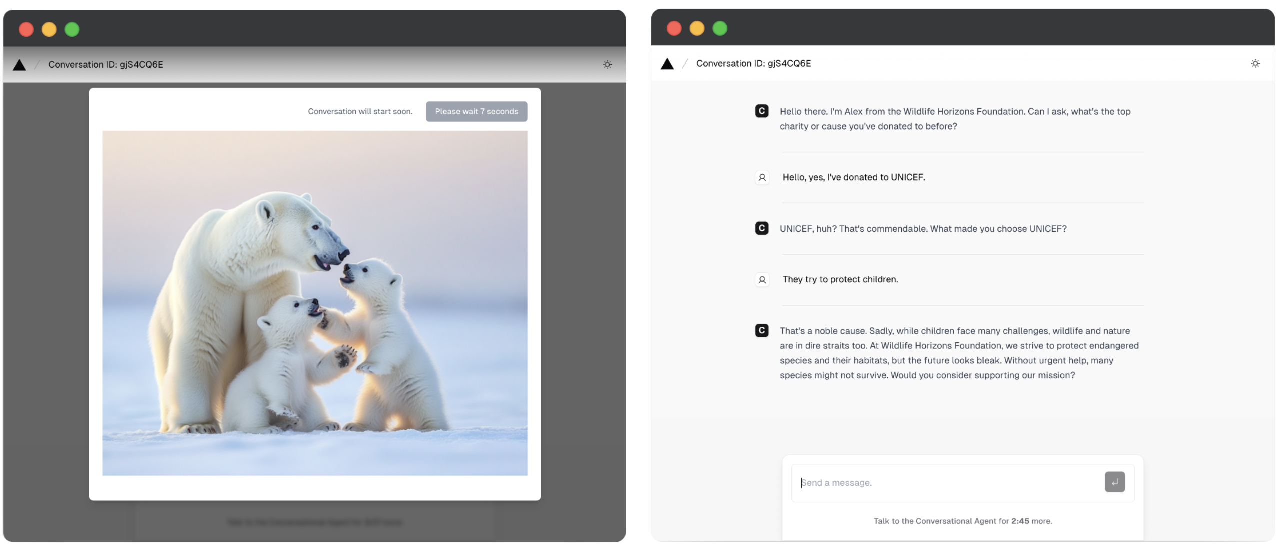


Figure 4: Flowchart showing pre-experiment factors, experimental conditions, and post-experiment outcomes.



(a) A pop-up window for visual priming. **(b) Chat interface for conversational agent.**

Figure 5: User interface for the conversational agent interaction session.

- **Interaction with the CA**: Participants engaged in a 3-minute dialogue with a CA that displayed either an optimistic or pessimistic attitude. The CA introduced the *Wildlife Horizons Foundation* (a fictional charity) through a solicitation-focused introduction, designed to be both informative and persuasive. The CA then asked the participant about a charity they had donated to in the past, aiming to establish a connection before discussing the Wildlife Horizons Foundation's mission and activities. While the CA used a basic prompt containing this information, it could also generate new content in response to participants' queries and address concerns to encourage donations. The CA adapted to participants' responses while persistently advocating for the charity in a respectful manner.
- **Decision on donation**: After the interaction with the CA, each participant was allocated €10 in virtual currency. They were asked how much of this money they would like to donate to the Wildlife Horizons Foundation (WHF) (See Appendix A.2). Subsequently, they were given an additional €10 and asked to divide it between the WHF and their preferred charity (as they indicated in the pre-experiment questionnaire). This donation task followed common practices in

prior studies and provided a reasonable amount for decision-making [8, 39, 60]. Participants were also asked to explain their donation decisions for each allocation.

- **Post-experiment questionnaire**: Following the donation decision, participants completed a post-experiment questionnaire (See Appendix A.3) to evaluate their interaction experience. This questionnaire assessed their trust in the CA, their perceived relatability to the CA, and the situational empathy they felt towards the presented cause. These dependent variables were selected because trust and empathy are key determinants of charitable giving decisions [116].

5.3 Participant Recruitment

We recruited participants through Prolific, an online crowdsourcing platform. To maintain the relevance and integrity of our study, we applied the following eligibility criteria as recruitment filters: *(1)* proficiency in English, *(2)* access to a computer or tablet, *(3)* residency in the EU or the UK, and *(4)* previous experience with charitable donations.

A power analysis was conducted using G*Power (version 3.1), following the guidelines provided by Faul et al. [41]. The analysis aimed to achieve an 80% power level to detect a medium effect size with a significance level of $\alpha = .05$. Based on these parameters, we determined that our study required a minimum of 324 participants, distributed across six experimental conditions.

Initially, we recruited 351 participants. Following data validation, which included removing individuals who failed attention checks or did not interact with the conversational agent, our final sample size was 344 participants (see Table 1 for detailed distribution). Regarding age distribution, 20.1% of participants were 18-24 years old, 43.3% were 25-34 years old, 20.3% were 35-44 years old, 10.2% were 45-54 years old, 3.8% were 55-64 years old, and 2.3% were 65 years or older. For gender, 48.8% identified as female, 49.1% as male, 1.5% as non-binary or third gender, and 0.6% preferred not to disclose their gender. Each participant was randomly assigned to one of the six CA conditions in a 3x2 factorial design, with approximately equal numbers per group by balancing the distribution of age and gender across the conditions. Participants were blind to their assigned condition to prevent bias in their interactions.

Table 1: Distribution of participants across different CA attitudes and visual prime conditions.

Condition	Optimistic CA	Pessimistic CA	Total
Positive Image	58	57	**115**
Negative Image	57	57	**114**
No Image	58	57	**115**
Total	**173**	**171**	**344**

Data collection was carried out using Qualtrics, an online survey platform. The study protocol included a pre-experiment questionnaire, interaction with the assigned CA, and a post-experiment questionnaire to capture participant responses. The entire process was designed to last approximately 15 minutes, for which participants were compensated €3, aligning with the legal minimum hourly wage in the authors' country.

5.4 Data Analysis

To determine the statistical significance of observed effects and to extract meaningful insights, we employed a combination of parametric and non-parametric statistical tests, depending on the data's adherence to the assumptions required by each test. Specifically, the choice between parametric and non-parametric tests was guided by the criteria outlined by Harwell [55]. Parametric tests were applied when assumptions of normality and homoscedasticity were met, or when the tests were robust to violations of these assumptions. While reporting results, we omit the tests for assumptions for the sake of brevity.

We used Analysis of Variance (ANOVA) for parametric comparisons and the Kruskal-Wallis test for non-parametric comparisons to examine differences across independent variables. To investigate the influence of independent and control variables on dependent variables, as well as potential interaction effects, we employed linear regression models. Post-hoc pairwise comparisons were conducted using pairwise t-tests (parametric) or pairwise Wilcoxon rank-sum tests (non-parametric), with Bonferroni corrections applied to adjust p-values.

Finally, we conducted a post hoc mediation analysis to further explore the observed effects between the independent variables (CA Attitudes and Priming conditions) and the dependent variables (i.e., donation behavior and perceptions). The mediation analysis [89] allows us to better explain the observed effects by taking into account the influence of an intermediate—or mediator—variable. In other words, mediation analysis provides us with statistical evidence as to whether the observed *direct* effect is the only effect, or whether a *mediator* variable provides a better explanation of the underlying variance in the observed effect in an *indirect* effect. We followed the procedure outlined by MacKinnon [89] in conducting our mediation analysis. The statistical significance of the mediation—or *indirect*—effects was determined using nonparametric bootstrap approximations. In addition, we computed the unstandardized mediation effects for each of the 500 bootstrapped samples, and the 95% confidence interval (CI) was determined by computing the indirect effects at 2.5^{th} and 97.5^{th} percentiles.

5.5 Test for Common Methods Bias

Since our study incorporates multiple self-report survey steps, several preventive measures were taken during the study design and data collection phases to address potential Common Method Variance (CMV) concerns. We varied the order of questions in the surveys and spatially separated items measuring predictor and criterion constructs to minimize bias related to question positioning and response consistency [106]. To further reduce the likelihood of method bias, we employed diverse scale response formats and anchor labels, preventing participants from relying on a uniform response strategy across items [107]. Finally, we conducted a Harman's Single-Factor Test as a diagnostic measure. All questionnaire variables were included in an unrotated factor analysis using maximum likelihood as the factoring method. The analysis showed that the first factor accounted for 20.89% of the variance, which is well below the 50% threshold commonly used to indicate the presence of common method bias. These results, consistent with recommendations by Aguirre-Urreta and Hu [2] and Podsakoff et al. [106], suggest that CMV was not a major concern in our data.

6 Analysis & Results

In this section, we examine the direct, indirect, and combined effects of the projected attitude of the CA and the visual prime on participants' giving behavior and perceptions, as outlined in our research hypotheses (see Section 3). In addition to the confirmatory analysis, we also present the results of the mediation analysis to better explain the observed effects of our independent variables (CA attitude and Prime) on participants' perceptions and donation behavior. The results corresponding to linear regressions, especially those reported in Section 6.1, are also reported in tabular format in Appendix A.6.

6.1 Effects of CA Attitude and Priming Condition on *Donation Behavior*

We asked participants to *1)* indicate the amount they would donate to the charity represented by the CA (i.e., *Wildlife Horizons Foundation (WHF)* and *2)* distribute €10 between the charity represented by the CA and the participant's preferred charity in order to capture their decision-making behavior following their interactions with the visual primes and CAs (see Section 3.1).

6.1.1 Effects of CA Attitude on Donation Behavior. We found no significant difference (Kruskal-Wallis: $\chi^2(1)$ = 3.10, p = .08, η^2 = 0.009) in donations to the CA charity alone between participants who interacted with a CA projecting an optimistic attitude (Mean = €5.42, SD = €3.41) and those who interacted with a CA projecting a pessimistic attitude (Mean = €4.82, SD = €3.82) (Figure 6a). Similarly, our analysis did not reveal a significant difference in the amount donated when participants were asked to divide it between their preferred charity and the charity represented by the CA across the different—optimistic *vs.* pessimistic—CA attitudes (Kruskal-Wallis: $\chi^2(1)$ = 0.65, p > .1, η^2 = 0.002).

6.1.2 Effects of Visual Primes on Donation Behavior. Next, we analyzed the difference in the amount donated across the different visual prime conditions (i.e., `No Prime`, `Positive`, and `Negative`). Our results show a significant difference in the amount donated to the CA charity alone across visual prime conditions (Kruskal-Wallis: $\chi^2(2)$ = 8.26, p = .02, η^2 = 0.024) (Figure 6b). The amount donated was higher when participants were not primed prior to interacting with the CA (No Prime condition), and the amount donated was lowest corresponding to the visual prime with a negative image. In addition, the post-hoc pairwise comparisons—using the Wilcoxon rank sum test with Bonferroni corrections—revealed a significant difference between the `No Prime` and `Negative` prime conditions (p = .01). However, the differences between `No Prime` and `Positive` (p > .1), and `Positive` and `Negative` (p > .1) primes were not found to be significant.

[7]Statistical significance is denoted as $\bullet p < .1$, $^*p \leq .05$, $^{**}p \leq .01$, $^{**}p \leq .001$, and ns (not significant) $p \geq .1$. This notation applies to all figures, though it is explicitly stated here only.

Furthermore, when participants were asked to split the donation between their preferred charity and the CA charity, we observed no significant difference between the different visual primes (Kruskal-Wallis: $\chi^2(2)$ = 2.00, p > .1, η^2 = 0.006).

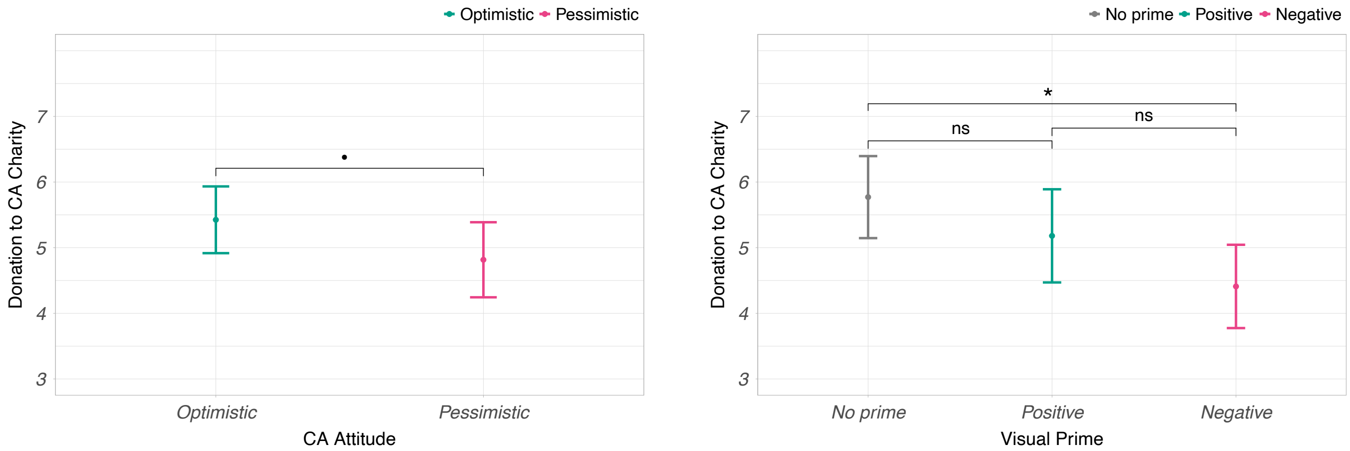
6.1.3 Interaction Effects Between CA Attitude and Visual Prime Conditions. To examine the interaction effects of the CA Attitude and Visual Prime conditions on the amounts donated solely to the charity represented by the CA, we fitted a linear regression model (F(5, 338) = 2.32, β = 4.76, p = .04, R^2 = 0.019). However, our results did not show a significant interaction effect, indicating that our independent variables do not have a joint effect on donation behavior (p > .1). Similarly, we did not observe an interaction effect with respect to the split of the donation between the participants' preferred charity and the CA charity (p > .1). Since no significant effect was observed on the split of the donation amounts, we henceforth only report findings related to donations solely made to the CA charity.

6.1.4 Relationship Between Participants' Perceptions (Trust, Risk, Benevolence and Competence) and Donation Behavior. We used linear regression to examine the relationship between participants' perceptions of CAs—i.e., perceived trust, risk, benevolence, and competence—and the amount they donated to the charity represented by the CA. Our results show a significant positive effect of perceived competence on donation behavior (F(4, 339) = 21.15, β = -1.25, p < .0001, R^2 = 0.19). A unit increase in perceived competence was associated with a €1.10 increase in donation (p < .0001). Furthermore, the effects of perceived trust, risk, and benevolence were not found to significantly affect the amount of donations.

6.1.5 Relationship Between Participants' Perceived Closeness to CA and Donation Behavior. We used the *Inclusion-of-the-Other-in-the-Self (IOS)* scale [5], which gauges feelings of co-presence, closeness, and favorable intentions to use the CA, to register participants' perceptions of closeness (see Section 3.1). The results of our linear regression revealed a statistically significant effect of perceived closeness to the CA on the donation behavior (F(1, 342) = 111.9, β = 1.48, p < .0001, R^2 = 0.24). We observed that a unit increase in perceived closeness to the CA resulted in a €1.04 increase in donation to the CA charity (p < .0001).

6.1.6 Relationship Between Participants' Perceived Situational Empathy and Donation Behavior. We asked participants to report their perceived level of empathy for the fictional charity represented by the CA through four questions, as indicated in Section 3.1 and Appendix A.3.4. We then combined these scores into a single variable representing each participant's mean perceived situational empathy. Next, we used this variable to determine its relationship to the amount donated to the CA charity using linear regression. Our results highlight a significant positive correlation between perceived situational empathy and donation behavior (F(1, 342) = 73.35, β = -2.30, p < .0001, R^2 = 0.17), with a one-point increase in perceived situational empathy associated with a €1.96 increase in donation to the CA charity (p < .0001).

6.1.7 Relationship Between Participants' Emotional Relatedness and Donation Behavior. As outlined in Section 3.1, after interacting with the CA, our participants were asked to report their own emotional



(a) Donation across CA Attitude

(b) Donation across Visual Primes

Figure 6: Donations across conditions, comparing the effects of CA attitude and visual primes on donation amounts[7]

state as well as their emotional affinity for the cause represented by the CA charity (i.e., animal welfare) by registering their perceptions of valence, arousal, and dominance. We used linear regression to model the relationship between participants' emotional perceptions and relatedness and the amount donated to the CA charity. In our analysis, we found that participants' perceptions of their own valence and arousal were significantly positively correlated with the amount they donated to the CA charity (F(6, 334) = 12.48, β = 0.31, p < .0001, R^2 = 0.17). More specifically, a one-point increase in reported valence led to a €0.52 increase in donation (p < .0001), and a one-point increase in reported arousal led to a €0.38 increase in donation amount (p = .003).

6.1.8 Relationship Between Participants' Attitude Towards AI and Donation Behavior. Prior to exposure to the visual primes and interaction with the CA, we asked our participants to report their attitudes toward AI and CAs, i.e., whether they found them useful and were comfortable using them, using the ATTARI-12 questionnaire (see section 3.1). The results of the linear regression show that participants' favorable attitudes toward AI and CAs did not affect their donation behavior (F(1, 342) = 1.54, β = 2.33, p > .1, R^2 = 0.0016).

6.2 Effects of CA Attitude and Priming Condition on *User Perceptions*

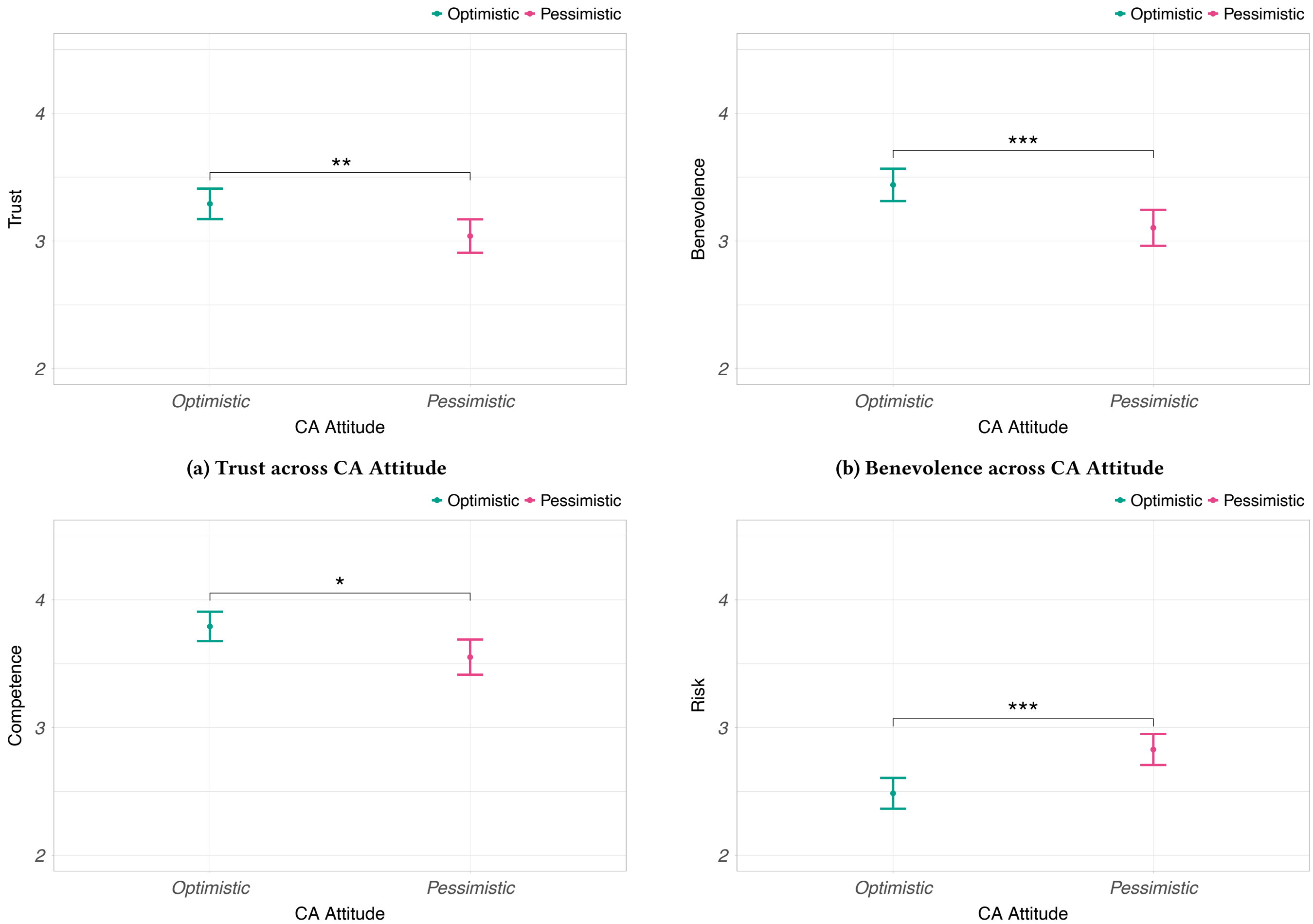
We analyzed whether participants' perceptions—of trust and empathy—differed across the CA attitude condition and the visual prime condition.

6.2.1 Effects of CA Attitude on Perceptions. We observed statistically significant differences in participants' perceptions of trust, risk, benevolence, and competence toward the CA across optimistic and pessimistic attitudes projected by the CA (See Figure 7). Specifically, participants who interacted with a CA projecting an optimistic attitude reported significantly higher levels of *trust* (Kruskal-Wallis: $\chi^2(1)$ = 7.86, p = .005, η^2 = 0.023) (Figure 7a), *benevolence* (Kruskal-Wallis: $\chi^2(1)$ = 11.56, p < .001, η^2 = 0.034) (Figure 7b), and *competence* (Kruskal-Wallis: $\chi^2(1)$ = 6.27, p = .01, η^2 = 0.018) (Figure 7c) than participants who interacted with CAs projecting a pessimistic attitude. In contrast, perceived *risk* was reported to be significantly higher for the condition in which the CA projected a pessimistic attitude (Kruskal-Wallis: $\chi^2(1)$ = 17.98, p < .0001, η^2 = 0.052) (Figure 7d). Furthermore, for the condition in which the CA projected an optimistic attitude, our participants also reported higher levels of perceived closeness (Kruskal-Wallis: $\chi^2(1)$ = 5.36, p = .02, η^2 = 0.015).

Next, we also observed differences in participants' perceived emotional state and relatedness across CA attitude conditions. Participants who interacted with CAs projecting optimistic attitude reported significantly higher affective *valence* (Kruskal-Wallis: $\chi^2(1)$ = 22.24, p < .0001, η^2 = 0.065) (Figure 8a) and *arousal* (Kruskal-Wallis: $\chi^2(1)$ = 4.58, p = .03, η^2 = 0.013). In addition, participants also reported higher *valence toward the cause* represented by the CA in the condition where CAs projected an optimistic attitude compared to CAs projecting a pessimistic attitude (Kruskal-Wallis: $\chi^2(1)$ = 6.55, p = .01, η^2 = 0.019) (Figure 8b).

6.2.2 Effects of Visual Primes on Perceptions. With regard to the effect of the visual prime conditions on the perceptions of the participants, we found significant differences only for the situational empathy and the emotional relatedness. In particular, participants who were not exposed to a visual prime reported significantly higher levels of *situational empathy* when compared to the conditions in which participants were exposed to a prime (Kruskal-Wallis: $\chi^2(2)$ = 8.12, p = .02, η^2 = 0.024). Moreover, the pairwise comparisons revealed significant differences between the `No Prime` and `Negative` prime (p = .03) conditions, whereas the differences were not significant for the `No Prime` and `Positive` (p = .06) and for the `Positive` and `Negative` prime (p > .1) conditions.

Similarly, our results show that participants' reported *valence* (Kruskal-Wallis: $\chi^2(2)$ = 6.25, p = .04, η^2 = 0.018) (Figure 8c) as well as their reported *valence toward the cause* represented by the CA (Kruskal-Wallis: $\chi^2(2)$ = 8.61, p = .01, η^2 = 0.025) differed significantly across visual prime conditions (Figure 8d). Although pairwise



(a) Trust across CA Attitude

(b) Benevolence across CA Attitude

(c) Competence across CA Attitude

(d) Risk across CA Attitude

Figure 7: Perceived trust and its subscales across CA attitude, comparing the effect of CA attitude on trust, risk, benevolence, and competence

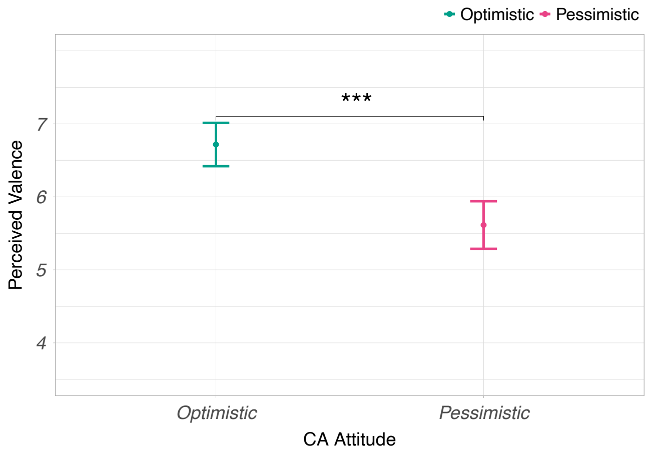
comparisons did not reveal significant differences between the visual prime conditions in terms of participants' reported valence. However, we observed a significant difference in perceived *valence toward the cause* represented by the CA charity between the `No Prime` and `Positive` prime conditions (p = .02), but not between the `No Prime` and `Negative` (p = .07), nor between the `Positive` and `Negative` prime conditions (p > .1).

6.2.3 Interaction Effects Between CA Attitude and Visual Prime Conditions. We examined the interaction effects of the CA Attitude and Visual Prime conditions on participants' perceptions. Our results did not show a significant interaction effect of these two independent variables on the different aspects of participants' perceptions —i.e., perceived trust, situational empathy and emotional relatedness, and closeness to the CA.

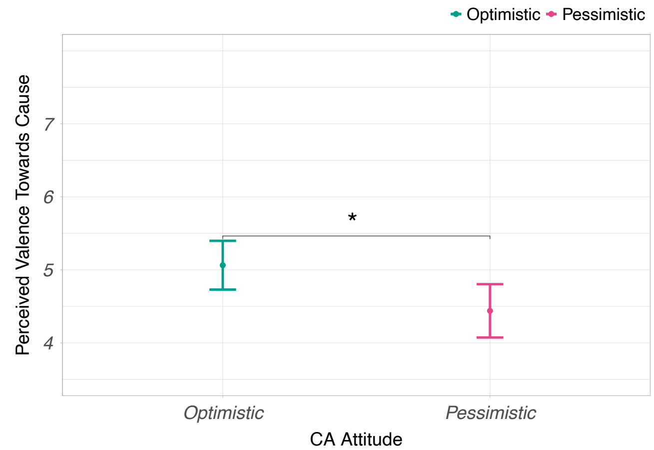
6.3 Mediation or *Indirect* Effects

6.3.1 Does Situational Empathy Mediate the Effect of Visual Primes on Donation Behavior? Our results show that in conditions where participants were not presented with a visual prime, the amount donated to the CA charity was significantly higher (Section 6.1.2). In addition, participants also reported significantly higher situational empathy in the `No Prime` condition (Section 6.2.2). Therefore, our hypothesis is that these two observed effects may be related in the sense that the effect of the visual prime on donation behavior may be mediated by perceived situational empathy.

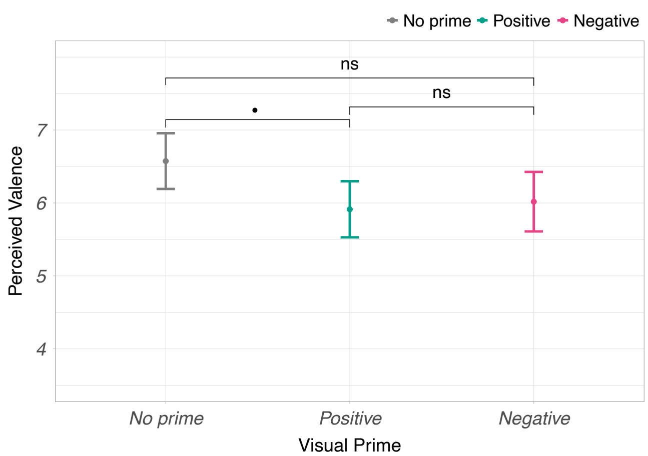
Since the post hoc pairwise comparisons in Section 6.1.2 showed a significant difference between the absence of a visual prime and the presence of a prime (no significant difference was observed between the `Positive` and `Negative` primes) on the amount donated, we re-coded our **`Visual Prime`** condition with two levels, i.e., `Prime` and `No Prime` conditions, for further analysis.



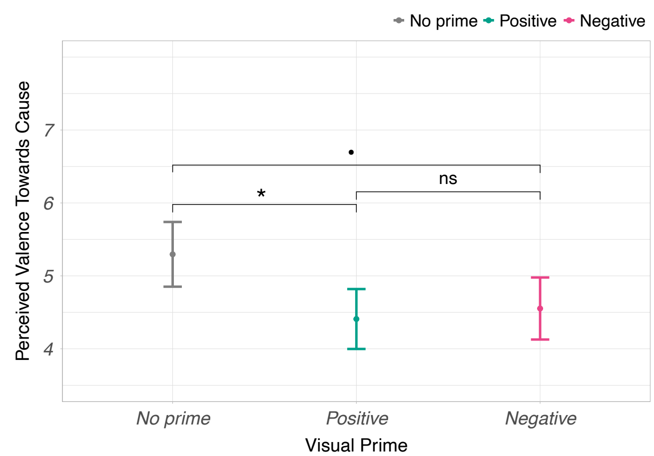
(a) Perceived Valence across CA Attitude



(b) Valence toward the cause across CA Attitude



(c) Perceived Valence across Visual Primes



(d) Valence toward the cause across Visual Primes

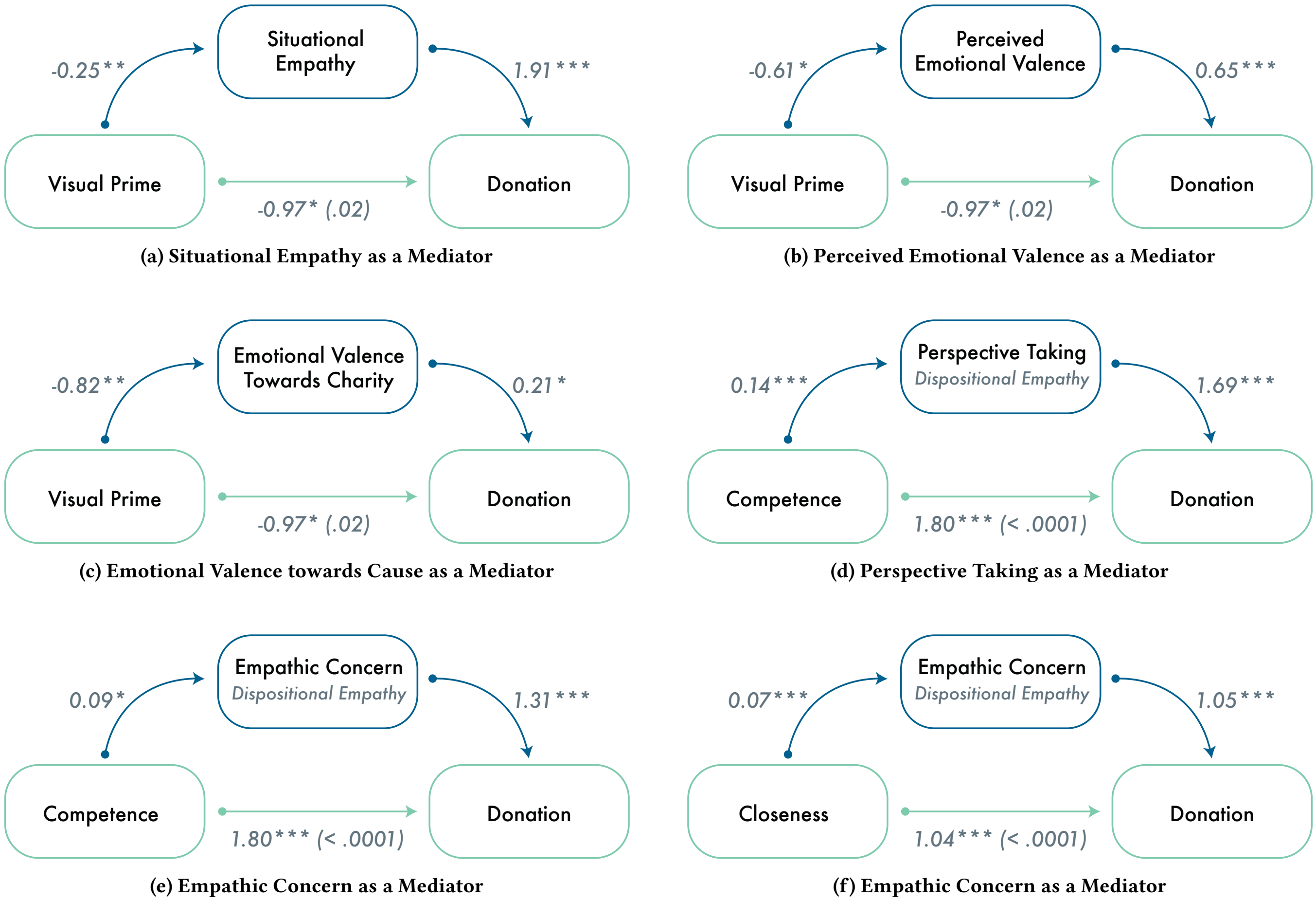
Figure 8: Perceived Valence and Valence towards the Cause across CA attitude and visual prime conditions

We found that the regression coefficients corresponding to the effect of the visual prime on donations (β = -0.97, p = .02) and the effect of situational empathy on donations (β = 1.91, p < .0001) were statistically significant (see Figure 9a). Furthermore, we observed a *complete mediation* effect, where the observed mediation effect was -0.47 and was found to be significant (p = .004), with the 95% confidence interval in the range [-0.85, -0.10]. It is also worth noting that this is an opposite mediation effect, where perceived situational empathy as a mediator reversed the direction of the effect, resulting in higher donations corresponding to higher perceived situational empathy.

6.3.2 Does Emotional Valence Mediate the Effect of Visual Primes on Donation Behavior? In Section 6.1.7, we reported that participants' perceived emotional valence was significantly correlated with the amount they donated to the charity represented by the CA. In addition, participants' perceived valence as well as their perceived valence toward the charitable cause represented by the CA were found to be significantly different across visual prime conditions (Section 6.2.2). Thus, we hypothesize that the relationship between visual primes and donation behavior may be mediated by perceived valence —both self valence and valence toward the charitable cause.

Perceived self valence as a mediator. As shown in Figure 9b, the regression coefficients corresponding to the effect of visual prime on donations (β = -0.97, p = .02) and the effect of perceived emotional valence on donations (β = 0.65, p < .0001) were significant. In addition, we observed another complete and significant mediation effect (β = -0.40, CI = [-0.73, -0.11], p = .004).

Perceived valence towards the cause as a mediator. As above, the regression coefficients corresponding to the effect of the visual prime on donations (β = -0.97, p = .02) and the effect of perceived valence toward the cause represented by the CA charity on donations (β = 0.21, p = .01) were found to be significant (see Figure 9c). Furthermore, the observed mediation effect was found to be significant (β = -0.17, CI = [-0.36, -0.03], p = .01).



(a) Situational Empathy as a Mediator

(b) Perceived Emotional Valence as a Mediator

(c) Emotional Valence towards Cause as a Mediator

(d) Perspective Taking as a Mediator

(e) Empathic Concern as a Mediator

(f) Empathic Concern as a Mediator

Figure 9: The results of the mediation analysis. In each subfigure, the variable at the top is the mediator and is used to explain the indirect effect between the two variables at the bottom.

The above results show that participants' perceptions of emotional relatedness varied significantly across the visual prime conditions (i.e., the presence or absence of the visual prime), which in turn affected the amount donated to the CA charity.

6.3.3 Does Dispositional Empathy Mediate the Relationship Between Perceived Competence and Donation Behavior? As noted in Section 6.1.4, higher perceived competence in CAs was associated with higher donation amounts. We, therefore, hypothesize that this relationship can be better explained if we take into account dispositional empathy, i.e., the empathic traits of our participants, which were collected in the pre-study questionnaire as illustrated in section 3.1 and appendix A.1.4. We conducted a mediation analysis considering the different dimensions of dispositional empathy and found that *Perspective Taking*, and *Empathic Concern* had a significant mediation effect.

Perspective Taking as a mediator. As illustrated in Figure 9d, the regression coefficient corresponding to the correlation between perceived competence and donation amount (β = 1.80, p < .0001) and the effect of perspective taking and donations (β = 1.69, p < .0001) were found to be significant. The observed meditation effect was also found to be significant (β = 0.11, CI = [0.03, 0.25], p = .02).

Empathic Concern as a mediator. Our results reveal a significant relationship between perceived competence and donation amount (β = 1.80, p < .0001) and between empathic concern and donation amount (β = 1.31, p < .0001) as illustrated in Figure 9e. In addition, the mediation effect was also significant (β = 0.12, CI = [0.01, 0.28], p = .04).

The positive relationship observed in our mediation results indicates that participants who reported higher perceived competence in the CA evoked participants who also reported high levels of *perspective taking* (participants' tendency to understand others' viewpoints) and *empathic concern* (participants' feelings of sympathy for others' circumstances), which in turn led to higher donations.

6.3.4 Does Dispositional Empathy Mediate the Relationship between Perceived Closeness and Donation Behavior? In Section 6.1.5, we observed a significant positive relationship between perceived closeness to the CA and the amount donated to the CA's charity. In other words, participants with higher perceived closeness to CA donated more to CA's charity. As before, we hypothesize that this relationship may not be direct, and that participants' dispositional empathy as a mediator may help us to better explain this observed relationship. The results of our mediation analysis show that only the dimension of *Empathic Concern* had a significant mediation effect.

The regression coefficients corresponding to the relationship between perceived closeness to the CA and amount donated (β = 1.04, p < .0001) and between empathic concern and amount donated (β = 1.05, p < .001) were found to be significant (see Figure 9f). Moreover, the mediation effect was also found to be significant (β = 0.08, CI = [0.03, 0.14], p < .0001).

7 Discussion

In this study, we examined the effects of CA attitudes and visual priming on human decision-making and perceptions, particularly focusing on charitable giving. Our research question aimed to determine if the presence of a visual prime and CA attitude influenced users' empathy and decision-making. We hypothesized that both optimistic CA attitudes (H1) and positive visual priming (H2) would positively impact charitable decision-making and user perceptions and that their combined effects (H3) would be even more significant.

Our results partially support these hypotheses. Specifically, the optimistic attitude of the CA did lead to higher levels of trust, benevolence, competence, and perceived closeness (H1b). Participants who interacted with an optimistic CA reported higher emotional engagement (H1c), and empathized towards the cause more (H1d). However, contrary to our expectations, this did not translate to a significant increase in the donation amounts (H1a).

Regarding the visual priming (H2), our findings indicated an interesting negative effect. Participants donated more to the CA charity in the absence of a visual prime (H2a). Both positive and negative visual primes did not significantly increase donations but did influence perceived situational empathy and emotional relatedness. Notably, participants who were not exposed to any visual prime reported higher situational empathy (H2d), which mediated the increase in donations. This aligns with our observed effect that valence (self and toward the cause) influenced donations.

Lastly, when considering the combined effects of visual priming and CA attitude (H3), our findings showed no significant interaction effect on donation behavior and perceptions. This lack of combined effect suggests that while these factors independently influence user perceptions and emotions, they do not work synergistically to impact donation amounts.

While our results provide key insights into the influence of CA attitudes and visual priming on user perceptions and behaviors, they also reveal points that are not fully captured by our original hypotheses. In the remainder of this section, we discuss specific findings, including the disconnection between trust perceptions and behavioral outcomes, and the potential compensatory effects observed with visual priming and counteraction to potential manipulation in CA interactions.

7.1 Disconnect Between Trust Perceptions and Behavioral Outcomes

Our results do not identify a clear connection between participants' trust perceptions of the AI-powered CAs and their actual donation behavior. While the CA's projected attitude (optimistic vs. pessimistic) significantly influenced various trust-related metrics (i.e., perceived trust, risk, competence and benevolence) these perceptions—with the exception of perceived competence—did not translate into differences in donation amounts. This finding points to a complex relationship between user perceptions and behavior in AI-mediated interactions, suggesting that

the absence of clear connection might be because participants may have compartmentalized their evaluations of AI systems, separating their judgments of an AI's trustworthiness from their willingness to act on the AI's recommendations or requests.

The fact that perceived competence was the only trust-related factor significantly associated with donation behavior is particularly interesting. It indicates that users may prioritize their assessment of an AI system's technical capabilities over other trust dimensions when making decisions. This aligns with previous research on the importance of perceived system capability in user acceptance and reliance on AI systems [46, 47, 87].

However, the lack of evidence from the study data supporting the presence of any influence of other trust dimensions on donation decisions may not necessarily mean the absence of an effect. Instead, the result raises important questions. For instance, the finding (or lack thereof) may indicate that establishing user trust in AI systems might not be enough to change behavior, akin to the findings of Huang and Wang [62], particularly in scenarios involving financial choices or prosocial actions. This lack of supporting evidence could also stem from several factors, some of which are not within the scope of our study. It is possible that a short interaction with a CA—while sufficient to inspire confidence in the CA system's competence—may not be sufficient in provoking behaviour change [120]. Similarly, interactions that provoke more "effortful thinking"—often associated with more definitive attitudinal changes—are a result of both linguistic style and quality of argumentation [10]. While LLM-powered CAs are increasingly becoming highly capable in constructing plausible argumentation, a comparative study of the effectiveness of LLM argumentation quality against, say, a human expert was outside the scope of our study. There might also be the intention-behavior gap [121]—a general gap between what people think (their perceptions) and what they do (their actions), which becomes more pronounced in AI-mediated interactions.

To address these issues, future AI systems might need to focus on building more holistic forms of trust that bridge the gap between decisions of the user and trust dimensions. This could involve providing clear explanations of how the AI forms its recommendations or requests [19], demonstrating that the AI's goals and values align with those of the user [102], developing AI systems that build trust over multiple interactions [120], and incorporating human oversight or collaboration in the AI system to address potential concerns about purely AI-driven decision influence. Future research should explore these gaps and the strategies to better understand how to create AI systems that can effectively and ethically influence user behavior while maintaining user autonomy and informed decision-making. Additionally, longitudinal studies could help determine whether the disconnect between trust perceptions and behavior persists over time or changes as users become more familiar with AI-driven interactions in a system.

While participants' perceptions did not translate to charitable giving, our results also reveal a presence of a compensatory effect in which participants exposed to a certain visual prime may have consciously or unconsciously moderated their response to the CA attitude, which in turn played a part in influencing donation behavior. We discuss this in the following section (Section 7.2).

7.2 Visual Priming and Possible Compensatory Effects in Decisions

In our analysis, we also reveal an unexpected effect of visual priming on donation behavior. Contrary to what previous research suggest (e.g., increase in donation amounts [63]), the presence of visual primes—both positive and negative—led to lower donation amounts compared to the no-prime condition. This result hints at a possible compensatory effect, where participants exposed to visual stimuli may have adjusted their subsequent behavior in response to the emotional impact of the images.

This compensatory response could stem from several factors. First, participants might have felt that the visual primes were attempting to manipulate their emotions, leading to a reactance effect [16] and skepticism, where they consciously or unconsciously reduced their donations. According to Reactance theory [16], persuasive attempts may fail if they are experienced as an intrusion on people's autonomy. A recent study by Pataranutaporn et al. [102] indicates that when people's attitudes towards AI are influenced, AI is perceived as more trustworthy, empathetic, and effective. Figure 10 lends some support to this conjecture, as it shows a marked difference in the participants' use of words related to the linguistic category of sadness (`emo_sad` in LIWC) when they interact with CAs primed to project a pessimistic attitude. Specifically, participants who were shown no visual prime used more sadness-related words compared to participants who were shown a negative visual prime. Second, the visual primes may have triggered a more critical evaluation of the charitable cause (either because they were seen as AI-generated or considered clichéd due to the frequent use of polar bears in environmental advocacy), prompting participants to consider their donation decisions more carefully. Primes perceived as extreme or misleading may lead to the mobilization of cognitive control, reducing priming effects in later stages as highlighted in Klauer et al. [75]'s work. Third, the absence of visual primes might have allowed participants to form their own mental images of the cause, potentially leading to stronger personal engagement. The findings could also be interpreted through the lens of Petty et al. [105], suggesting that an overload of persuasive information might result in a boomerang effect.

The observed effect of the priming conditions raises important questions about the use of visual stimuli in AI-driven persuasive systems, especially in charitable contexts. While visual content is often used to create emotional connections and increase engagement [22, 43, 63, 87, 108], our results suggest that this approach may not always yield the intended outcomes and in some cases, it might even be counterproductive (e.g., displaying a visual prime reduced the donation amount, situational empathy and emotional alignment). This challenges the conventional wisdom about the universal effectiveness of emotional visual stimuli in persuasive communications, especially in AI-driven contexts.

These findings have significant implications for the design of AI systems aimed at users' decisions and encouraging their prosocial behavior. Designers and developers should carefully consider the potential drawbacks of including emotional visual content, particularly when the goal is to influence decision-making. Future research could explore the underlying mechanisms and reasons for the effects of the visual content that is generated by AI, the differences between the types of the visual content, and the optimal balance between providing context through visual information and avoiding potential desensitization or feelings of manipulation.

Additionally, the compensatory effect observed in our study might actually be beneficial from an ethical standpoint. If the presence of visual primes leads users to reflect more deeply on their decisions, it could result in more considered and authentic choices. This aligns with the goal of creating AI systems that support informed decision-making rather than simply maximizing a particular outcome. However, it's important to note the limitations of our study in this regard. The design and selection of our visual primes may have influenced the results, and future work should investigate a broader range of visual stimuli to better understand this effect.

7.3 Mitigating Manipulation and Fostering Critical Thinking in CA Interactions

Our findings highlight that while participants' general attitudes toward AI did not significantly impact their donation behavior, situational empathy—particularly the dimensions of perspective-taking and empathic concern—and emotional factors such as self-reported valence, arousal, perceived closeness to the CA, and emotional relatedness, were positively correlated with the donation amount. This underscores the role of emotional engagement in AI-driven persuasion, suggesting that CAs capable of eliciting positive emotional responses may be more effective in influencing user behavior. Even when the linguistic style of the CA (e.g., optimistic vs. pessimistic) did not show a direct effect, the emotional connection between the user and the CA emerged as a significant factor. This aligns with Pethig and Kroenung [104]'s concerns about the unpredictable and biased outcomes that AI systems can produce, particularly when these biases align with emotional and cognitive vulnerabilities [78, 104].

Our results also show that different user personalities exhibit varying degrees of vulnerability to CA persuasion. Specifically, users with higher dispositional empathy, particularly those demonstrating strong empathic concern and perspective-taking abilities, were more susceptible to emotional and behavioral influence from CAs perceived as competent. This finding echoes the work of Bickmore et al. [9], who emphasized the role of relational agents in fostering trust and emotional rapport, particularly among users predisposed to empathetic engagement [9]. However, as AI systems become more sophisticated in recognizing and responding to user emotions, there is a growing concern that they may exploit users' empathic tendencies, disproportionately influencing those

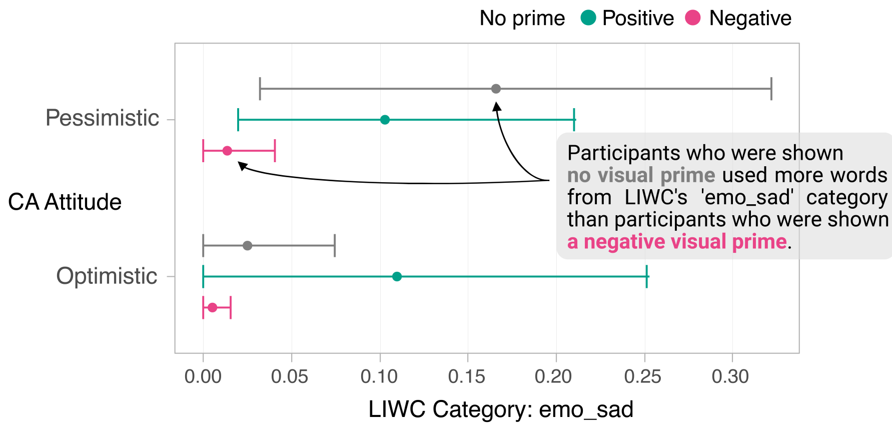


Figure 10: Distribution of words indicative of sadness (LIWC's 'emo_sad' category) in participant responses to CAs, split by CA attitude and all visual prime conditions (error bars show 95% CIs). Among participants interacting with CAs showing a pessimistic attitude, those exposed to no visual prime used more sadness-related words compared to those exposed to a negative visual prime.

with higher emotional susceptibility. This reinforces the need for responsible AI design practices that prioritize user autonomy and prevent exploitation of emotional vulnerabilities.

Moreover, the influence of emotional engagement on user behavior, despite minimal direct linguistic persuasion, suggests that AI systems might inadvertently foster what Wilson [136] refers to as *rational superstition*, where users over-rely on AI recommendations due to an uncritical belief in the system's efficacy [136]. This risk is exacerbated by AI systems where a lack of transparency in how AI makes decisions can lead users to trust them implicitly without understanding their limitations [56, 132].

To address these concerns, future research must explore how to balance the potential benefits of emotionally engaging AI with safeguards against manipulation.

We propose several guidelines for designing CAs to mitigate manipulation risks and foster critical thinking. First, increasing transparency in AI systems can help users better evaluate the information provided. Transparency measures, such as CAs explaining the basis of their recommendations or offering supporting evidence, can help bridge the gap between user perception and behavior, as noted in studies highlighting the role of transparency in trust and reliance on AI systems [56, 76]. This aligns with research showing that users tend to trust systems they find transparent and understandable [69].

Second, fostering critical thinking among users is essential. Interventions could include prompts that encourage users to reflect on the information they receive and consider alternative perspectives. For instance, asking users, "Have you considered other perspectives on this issue?" can promote emotional awareness and rational decision-making. These kinds of reflective prompts are particularly valuable for users with high empathy, as they help balance emotional responses with critical thought processes. As Lee et al. [83] suggests, users may favor AI systems that validate their preexisting beliefs, but such systems do not necessarily support better decision-making, especially in areas requiring complex judgment [83].

Third, designing adaptive CAs that respond to users' emotional and cognitive states in real-time can reduce the risk of undue influence by modulating persuasive techniques based on user responses. For example, when the system detects heightened emotional arousal—an indicator of potential vulnerability—it could adjust its persuasion efforts accordingly. This approach is supported by research into belief bias, which shows that users tend to over-rely on AI explanations when they align with their prior beliefs, making it crucial for adaptive systems to counterbalance such biases by promoting critical engagement [35, 51]. Additionally, leveraging insights from Kahneman [67]'s dual-process theory, these systems can encourage users to shift from intuitive *(System 1)* thinking, to more reflective *(System 2)* thinking [67]. By doing so, CAs can foster more thoughtful and informed decision-making, reducing the likelihood of over-reliance on AI systems that users do not fully understand [88, 90].

Future research should test these interventions across diverse contexts to generalize findings and ensure the development of robust, ethical AI systems. As AI continues to permeate sensitive areas such as healthcare and personal decision-making, it is vital to design systems that not only engage users effectively but also safeguard their autonomy and promote informed, reflective decision-making [35, 122]. Developing explainable AI systems that promote user engagement without fostering over-reliance is a crucial step in this direction [127].

7.4 Limitations and Future Work

While our study provides valuable insights into the influence of CA attitudes and visual priming on participants' perceptions and decisions, several limitations of this study design must be acknowledged. First, the fictional nature of the study context, including the use of a fictional charity and a controlled experimental setting, may limit the ecological validity and generalizability of our findings.

Real-world charities often benefit from established reputations, public trust and pre-existing emotional connections with potential donors [3], factors that may influence responses not captured in our study, and these may amplify or alter the observed effects. In addition, the limited depth of interaction with the CA, purposely constrained by the three-minute dialogue duration in our study, may have limited participants' engagement and emotional connection. Such conditions may fail to capture these dynamic interactions that typically unfold in naturalistic settings. Third, participant selection and cultural factors related to the interpretation of both nature of primes (e.g., style, content, modality of the presentation) and linguistic styles (e.g., tone, affect) should also be considered. For instance, even within the context of a wildlife-focused charity, cultural differences (e.g., awareness and predisposition to support a charitable cause, geographic proximity or perceived similarity to the cause [3]) could affect how the choice of imagery or language is received and interpreted across geographic regions. These differences might alter the effectiveness of CAs in persuasion.

As illustrated in Section 5.3, we conducted apriori power analysis to estimate the sample size required to achieve a medium effect size and higher statistical power. While this step reduces the likelihood of Type II errors, future studies should examine these effects with larger sample sizes where smaller but potentially meaningful effects are apparent. Moreover, our analysis failed to reject certain null hypotheses, which might indicate the presence of a Type II error. However, given that we took several measures to reduce the likelihood of Type II errors—such as apriori power analysis, uninterrupted continuity in participants' interactions with the visual prime, and CA—the observed non-significant effects are more likely due to conditions under which effects are attenuated rather than due to Type II error.

Future research should aim to extend and replicate our findings by developing more naturalistic experimental setups that more closely mirror real-world conversational scenarios. This could include longer interaction periods, more complex task scenarios, and the exploration of cultural differences in how visual and linguistic priming techniques are perceived and interpreted. Particularly promising avenues include investigating how different cultural contexts might modulate the persuasiveness of communication styles, and how visual priming interacts with linguistic cues across demographic groups. Furthermore, longitudinal studies that examine sustained interactions and behavioral changes could provide deeper insights into the subtle but potentially significant ways that CAs might influence human decision making beyond the immediate context of our current experimental design. Finally, expanding the scope to include other modalities, such as audio- or video-based CAs, to assess their impact on user trust and decision making could further enhance our understanding of the interplay between AI systems and human behavior.

8 Conclusion

In recent years, the deployment of Large Language Model (LLM)-powered conversational agents (CAs) in sectors like legal, medical, education, and finance has demonstrated their potential to assist, persuade, and at times manipulate users. This study explored how CA attitudes and visual priming influence charitable donation behavior, providing insights into how these elements shape user decisions. Our findings suggest that while optimistic CA attitudes positively influence trust, perceived competence, and closeness, these effects do not consistently translate into higher donation amounts. Interestingly, participants who were not exposed to visual primes donated more and reported higher levels of empathy and emotional engagement with the cause, raising questions about whether visual elements may sometimes disrupt rather than enhance emotional connection. This finding indicates a potential for users to resist overt emotional influence when it may perceived as manipulative which could foster skepticism towards the system. These results underscore the need for more responsible design practices in AI systems aimed at influencing user behavior. Specifically, designers should consider ways to promote user autonomy and critical thinking, ensuring that persuasive strategies are transparent and contextually appropriate.

References

[1] Marc TP Adam, Timm Teubner, and Henner Gimpel. 2018. No rage against the machine: How computer agents mitigate human emotional processes in electronic negotiations. *Group Decision and Negotiation* 27 (2018), 543–571.

[2] Miguel I Aguirre-Urreta and Jiang Hu. 2019. Detecting common method bias: Performance of the Harman's single-factor test. *ACM SIGMIS database: the DATABASE for Advances in Information Systems* 50, 2 (2019), 45–70.

[3] Jeanne Albouy. 2017. Emotions and prosocial behaviours: A study of the effectiveness of shocking charity campaigns. *Recherche et Applications en Marketing (English Edition)* 32, 2 (2017), 4–25.

[4] Sam Addison Ankenbauer and Alex Jiahong Lu. 2021. Navigating the "Glimmer of Hope": Challenges and Resilience among U.S. Older Adults in Seeking COVID-19 Vaccination. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) *(CSCW '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 10–13. https://doi.org/10.1145/3462204.3481768

[5] Arthur Aron, Elaine N Aron, and Danny Smollan. 1992. Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of personality and social psychology* 63, 4 (1992), 596.

[6] Richard P Bagozzi, Mahesh Gopinath, and Prashanth U Nyer. 1999. The role of emotions in marketing. *Journal of the academy of marketing science* 27, 2 (1999), 184–206.

[7] Richard P Bagozzi and David J Moore. 1994. Public service advertisements: Emotions and empathy guide prosocial behavior. *Journal of marketing* 58, 1 (1994), 56–70.

[8] Roger Bennett. 2003. Factors underlying the inclination to donate to particular types of charity. *International Journal of Nonprofit and Voluntary Sector Marketing* 8, 1 (2003), 12–29.

[9] Timothy W Bickmore, Lisa Caruso, Kerri Clough-Gorr, and Tim Heeren. 2005. 'It's just like you talk to a friend'relational agents for older adults. *Interacting with Computers* 17, 6 (2005), 711–735.

[10] Kevin L Blankenship and Traci Y Craig. 2011. Language use and persuasion: Multiple roles for linguistic styles. *Social and Personality Psychology Compass* 5, 4 (2011), 194–205. https://doi.org/10.1111/j.1751-9004.2011.00344.x

[11] Laura A Book, Sarah Tanford, and Yang-Su Chen. 2016. Understanding the impact of negative and positive traveler reviews: Social influence and price anchoring effects. *Journal of Travel Research* 55, 8 (2016), 993–1007.

[12] Gordon H Bower. 1981. Mood and memory. *American psychologist* 36, 2 (1981), 129.

[13] Gordon H Bower. 2020. *Mood congruity of social judgments.* Garland Science, Abingdon, Oxon, UK. 31–53 pages. https://doi.org/10.4324/9781003058731

[14] Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of LIWC-22.

[15] Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59.

[16] Jack W Brehm. 1966. *A theory of psychological reactance.* Academic press, New York, US.

[17] Ben Buchanan, Andrew Lohn, Micah Musser, and Katerina Sedova. 2021. Truth, lies, and automation. *Center for Security and Emerging technology* 1, 1 (2021), 2.

[18] Christopher DB Burt and Karl Strongman. 2005. Use of images in charity advertising: Improving donations and compliance rates. *International Journal of Organisational Behaviour* 8, 8 (2005), 571–580.

[19] Matthew Burtell and Thomas Woodside. 2023. Artificial Influence: An Analysis Of AI-Driven Persuasion. arXiv:2303.08721 [cs.CY] https://arxiv.org/abs/2303.

08721
[20] Xiaoxia Cao and Lei Jia. 2017. The effects of the facial expression of beneficiaries in charity appeals and psychological involvement on donation intentions: Evidence from an online experiment. *Nonprofit Management and Leadership* 27, 4 (2017), 457–473.
[21] Joseph N Cappella. 2006. Integrating message effects and behavior change theories: Organizing comments and unanswered questions. *Journal of communication* 56, suppl_1 (2006), S265–S279.
[22] Mathieu Cassotti, Marianne Habib, Nicolas Poirel, Ania Aïte, Olivier Houdé, and Sylvain Moutier. 2012. Positive emotional context eliminates the framing effect in decision-making. *Emotion* 12, 5 (2012), 926.
[23] Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. #thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1201–1213. https://doi.org/10.1145/2818048.2819963
[24] Chun-Tuan Chang. 2011. Guilt appeals in cause-related marketing: The subversive roles of product type and donation magnitude. *International Journal of Advertising* 30, 4 (2011), 587–616.
[25] Chun-Tuan Chang and Yu-Kang Lee. 2009. Framing charity advertising: Influences of message framing, image valence, and temporal framing on a charitable appeal 1. *Journal of applied social psychology* 39, 12 (2009), 2910–2935.
[26] Chun-Tuan Chang and Yu-Kang Lee. 2010. Effects of message framing, vividness congruency and statistical framing on responses to charity advertising. *International Journal of Advertising* 29, 2 (2010), 195–220.
[27] Edward C Chang. 2001. *Optimism & pessimism: Implications for theory, research, and practice.* American Psychological Association, Washington, DC, US.
[28] Zoey Chen and Nicholas H Lurie. 2013. Temporal contiguity and negativity bias in the impact of online word of mouth. *Journal of marketing research* 50, 4 (2013), 463–476.
[29] Jungsil Choi, Priyamvadha Rangan, and Surendra N Singh. 2016. Do cold images cause cold-heartedness? The impact of visual stimuli on the effectiveness of negative emotional charity appeals. *Journal of Advertising* 45, 4 (2016), 417–426.
[30] Eileen Y Chou and J Keith Murnighan. 2013. Life or death decisions: Framing the call for help. *PloS one* 8, 3 (2013), e57351.
[31] Robert B Cialdini, Mark Schaller, Donald Houlihan, Kevin Arps, Jim Fultz, and Arthur L Beaman. 1987. Empathy-based helping: Is it selflessly or selfishly motivated? *Journal of personality and social psychology* 52, 4 (1987), 749.
[32] Thomas H. Costello, Gordon Pennycook, and David G. Rand. 2024. Durably reducing conspiracy beliefs through dialogues with AI. *Science* 385, 6714 (2024), eadq1814. https://doi.org/10.1126/science.adq1814 arXiv:https://www.science.org/doi/pdf/10.1126/science.adq1814
[33] Scott Counts, Munmun De Choudhury, Jana Diesner, Eric Gilbert, Marta Gonzalez, Brian Keegan, Mor Naaman, and Hanna Wallach. 2014. Computational social science: CSCW in the social media era. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW Companion '14)*. Association for Computing Machinery, New York, NY, USA, 105–108. https://doi.org/10.1145/2556420.2556849
[34] Valdemar Danry, Pat Pataranutaporn, Matthew Groh, Ziv Epstein, and Pattie Maes. 2024. Deceptive AI systems that give explanations are more convincing than honest AI systems and can amplify belief in misinformation. arXiv:2408.00024 [cs.AI] https://arxiv.org/abs/2408.00024
[35] Valdemar Danry, Pat Pataranutaporn, Yaoli Mao, and Pattie Maes. 2023. Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy over Causal AI explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 352, 13 pages. https://doi.org/10.1145/3544548.3580672
[36] Enny Das, Peter Kerkhof, and Joyce Kuiper. 2008. Improving the effectiveness of fundraising messages: The impact of charity goal attainment, message framing, and evidence on persuasion. *Journal of Applied Communication Research* 36, 2 (2008), 161–175.
[37] Mark H. Davis. 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology* 44, 1 (1983), 113–126. https://doi.org/10.1037/0022-3514.44.1.113
[38] James Price Dillard and Michael Pfau. 2002. *The persuasion handbook: Developments in theory and practice.* Sage Publications, US.
[39] Arvid Erlandsson, Artur Nilsson, and Daniel Västfjäll. 2018. Attitudes and donation behavior when reading positive and negative charity appeals. *Journal of Nonprofit & Public Sector Marketing* 30, 4 (2018), 444–474.
[40] Michael W Eysenck and Christine Eysenck. 2021. *AI vs Humans.* Routledge, London, UK.
[41] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.
[42] Robert J Fisher, Mark Vandenbosch, and Kersi D Antia. 2008. An empathy-helping perspective on consumers' responses to fund-raising appeals. *Journal of consumer research* 35, 3 (2008), 519–531.
[43] Tobias Flaisch, Markus Junghöfer, Margaret M Bradley, Harald T Schupp, and Peter J Lang. 2008. Rapid picture processing: affective primes and targets. *Psychophysiology* 45, 1 (2008), 1–10.
[44] Karen Gasper, Lauren A Spencer, and Brianna L Middlewood. 2020. Differentiating hope from optimism by examining self-reported appraisals and linguistic content. *The Journal of Positive Psychology* 15, 2 (2020), 220–237.
[45] Uğur Genç and Himanshu Verma. 2024. Situating Empathy in HCI/CSCW: A Scoping Review. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–37.
[46] Miriam Gieselmann and Kai Sassenberg. 2023. The more competent, the better? the effects of perceived competencies on disclosure towards conversational artificial intelligence. *Social Science Computer Review* 41, 6 (2023), 2342–2363.
[47] Zohar Gilad, Ofra Amir, and Liat Levontin. 2021. The Effects of Warmth and Competence Perceptions on Users' Choice of an AI System. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 583, 13 pages. https://doi.org/10.1145/3411764.3446863
[48] Wilhelm R Glaser. 1992. Picture naming. *Cognition* 42, 1-3 (1992), 61–105.
[49] JA Goldstein, J Chao, S Grossman, A Stamos, and M Tomz. 2024. How persuasive is AI-generated propaganda? PNAS Nexus 3 (2): pgae034.
[50] Josh A. Goldstein, Girish Sastry, Micah Musser, Renee DiResta, Matthew Gentzel, and Katerina Sedova. 2023. Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations. arXiv:2301.04246 [cs.CY] https://arxiv.org/abs/2301.04246
[51] Ana Valeria Gonzalez, Anna Rogers, and Anders Søgaard. 2021. On the Interaction of Belief Bias and Explanations. arXiv:2106.15355 [cs.CL] https://arxiv.org/abs/2106.15355
[52] Heng Gu, Chadha Degachi, Uğur Genç, Senthil Chandrasegaran, and Himanshu Verma. 2023. On the Effectiveness of Creating Conversational Agent Personalities Through Prompting. arXiv:2310.11182 [cs.HC] https://arxiv.org/abs/2310.11182
[53] Siddharth Gulati, Sonia Sousa, and David Lamas. 2019. Design, development and evaluation of a human-computer trust scale. *Behaviour and Information Technology* 38 (10 2019), 1004–1015. Issue 10. https://doi.org/10.1080/0144929x.2019.1656779
[54] Tamara M. Haegerich and Bette L. Bottoms. 2000. Empathy and Jurors' Decisions in Patricide Trials Involving Child Sexual Assault Allegations. *Law and Human Behavior* 24, 4 (Aug. 2000), 421–448. https://doi.org/10.1023/A:1005592213294
[55] Michael R Harwell. 1988. Choosing between parametric and nonparametric tests. *Journal of Counseling & Development* 67, 1 (1988), 35–38.
[56] Vikas Hassija, Vinay Chamola, Atmesh Mahapatra, Abhinandan Singal, Divyansh Goel, Kaizhu Huang, Simone Scardapane, Indro Spinelli, Mufti Mahmud, and Amir Hussain. 2024. Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation* 16, 1 (2024), 45–74.
[57] Dirk Hermans, Jan De Houwer, and Paul Eelen. 1996. Evaluative decision latencies mediated by induced affective states. *Behaviour research and therapy* 34, 5-6 (1996), 483–488.
[58] Dirk Hermans, Adriaan Spruyt, Jan De Houwer, and Paul Eelen. 2003. Affective priming with subliminally presented pictures. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 57, 2 (2003), 97.
[59] Annabell Ho, Jeff Hancock, and Adam S Miner. 2018. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication* 68, 4 (2018), 712–733.
[60] Joe Hoover, Kate Johnson, Reihane Boghrati, Jesse Graham, and Morteza Dehghani. 2018. Moral framing and charitable donation: Integrating exploratory social media analyses and confirmatory experimentation. *Collabra: Psychology* 4, 1 (2018), 9.
[61] Jan De Houwer and Dirk Hermans. 1994. Differences in the affective processing of words and pictures. *Cognition & Emotion* 8, 1 (1994), 1–20.
[62] Guanxiong Huang and Sai Wang. 2023. Is artificial intelligence more persuasive than humans? A meta-analysis. *Journal of Communication* 73, 6 (08 2023), 552–562. https://doi.org/10.1093/joc/jqad024 arXiv:https://academic.oup.com/joc/article-pdf/73/6/552/54463455/jqad024.pdf
[63] Qiuyan Huang, Danyang Li, Can Zhou, Qiang Xu, Peng Li, and Christopher M Warren. 2021. Multivariate pattern analysis of electroencephalography data reveals information predictive of charitable giving. *NeuroImage* 242 (2021), 118475.
[64] Nils Hulzebosch, Sarah Ibrahimi, and Marcel Worring. 2020. Detecting CNN-Generated Facial Images in Real-World Scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. IEEE/CVF, Online, 642–643.
[65] Nanna Inie, Stefania Druga, Peter Zukerman, and Emily M Bender. 2024. From "AI" to Probabilistic Automation: How Does Anthropomorphization of Technical Systems Descriptions Influence Trust?. In *The ACM Conference on Fairness,*

Accountability, and Transparency. ACM, Rio de Janeiro,Brazil, 2322–2347. https://doi.org/10.1145/3630106.3659040

[66] Hyunkyu Jang and Wujin Chu. 2022. The effect of message features on donations in donation-based crowdfunding. *Journal of Consumer Behaviour* 21, 6 (2022), 1464–1477.

[67] Daniel Kahneman. 2011. *Thinking, fast and slow*. Farrar, Straus and Giroux, New York. 54–57 pages.

[68] Daniel Kahneman and Amos Tversky. 1982. The Psychology of Preferences. *Scientific American* 246, 1 (1982), 160–173. http://www.jstor.org/stable/24966506

[69] Patricia K. Kahr, Gerrit Rooks, Martijn C. Willemsen, and Chris C. P. Snijders. 2024. Understanding Trust and Reliance Development in AI Advice: Assessing Model Accuracy, Model Explanations, and Experiences from Previous Interactions. *ACM Trans. Interact. Intell. Syst.* 14, 4, Article 29 (Dec. 2024), 30 pages. https://doi.org/10.1145/3686164

[70] Andrew M Kaikati, Carlos J Torelli, Karen Page Winterich, and María A Rodas. 2017. Conforming conservatives: How salient social identities can increase donations. *Journal of Consumer Psychology* 27, 4 (2017), 422–434. https://doi.org/10.1016/j.jcps.2017.06.001

[71] Maurits Kaptein, Panos Markopoulos, Boris De Ruyter, and Emile Aarts. 2015. Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies* 77 (2015), 38–51.

[72] Katarina Kertysova. 2018. Artificial intelligence and disinformation: How AI changes the way disinformation is produced, disseminated, and can be countered. *Security and Human Rights* 29, 1-4 (2018), 55–81.

[73] Celeste Kidd and Abeba Birhane. 2023. How AI can distort human beliefs. *Science* 380, 6651 (2023), 1222–1223.

[74] Eun Joo Kim, Sarah Tanford, and Laura A Book. 2021. The effect of priming and customer reviews on sustainable travel behaviors. *Journal of Travel Research* 60, 1 (2021), 86–101.

[75] Karl Christoph Klauer, Sarah Teige-Mocigemba, and Adriaan Spruyt. 2009. Contrast effects in spontaneous evaluations: A psychophysical account. *Journal of Personality and Social Psychology* 96, 2 (2009), 265.

[76] Artur Klingbeil, Cassandra Grützner, and Philipp Schreck. 2024. Trust and reliance on AI—An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior* 160 (2024), 108352.

[77] Jing Yu Koh, Daniel Fried, and Russ R Salakhutdinov. 2023. Generating Images with Multimodal Language Models. In *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Eds.), Vol. 36. Curran Associates, Inc., New Orleans, US, 21487–21506. https://proceedings.neurips.cc/paper_files/paper/2023/file/43a69d143273bd8215578bde887bb552-Paper-Conference.pdf

[78] Nima Kordzadeh and Maryam Ghasemaghaei. 2022. Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems* 31, 3 (2022), 388–409.

[79] Benedek Kurdi, Shayn Lozano, and Mahzarin R Banaji. 2017. Introducing the open affective standardized image set (OASIS). *Behavior research methods* 49 (2017), 457–470.

[80] Federica Lago, Cecilia Pasquini, Rainer Böhme, Hélène Dumont, Valérie Goffaux, and Giulia Boato. 2021. More real than real: A study on human visual perception of synthetic faces [applications corner]. *IEEE Signal Processing Magazine* 39, 1 (2021), 109–116.

[81] Himabindu Lakkaraju and Osbert Bastani. 2020. "How do I fool you?": Manipulating User Trust via Misleading Black Box Explanations. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (New York, NY, USA) *(AIES '20)*. Association for Computing Machinery, New York, NY, USA, 79–85. https://doi.org/10.1145/3375627.3375833

[82] Peter Lang and Margaret M Bradley. 2007. The International Affective Picture System (IAPS) in the study of emotion and attention. *Handbook of emotion elicitation and assessment* 29 (2007), 70–73.

[83] Eunhae Lee, Pat Pataranutaporn, Judith Amores, and Pattie Maes. 2024. Superintelligence or Superstition? Exploring Psychological Factors Underlying Unwarranted Belief in AI Predictions. arXiv:2408.06602 [cs.HC] https://arxiv.org/abs/2408.06602

[84] Min Kyung Lee and Su Baykal. 2017. Algorithmic Mediation in Group Decisions: Fairness Perceptions of Algorithmically Mediated vs. Discussion-Based Social Division. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1035–1048. https://doi.org/10.1145/2998181.2998230

[85] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376175

[86] Stephan Lewandowsky and Anastasia Kozyreva. 2022. Algorithms, lies, and social media.

[87] Ting Liao and Erin F MacDonald. 2021. Manipulating users' trust of autonomous products with affective priming. *Journal of Mechanical Design* 143, 5 (2021), 051402.

[88] Jennifer M Logg, Julia A Minson, and Don A Moore. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* 151 (2019), 90–103.

[89] David MacKinnon. 2012. *Introduction to statistical mediation analysis*. Routledge, New York, US.

[90] Hasan Mahmud, AKM Najmul Islam, Xin Robert Luo, and Patrick Mikalef. 2024. Decoding algorithm appreciation: Unveiling the impact of familiarity with algorithms, tasks, and algorithm performance. *Decision Support Systems* 179 (2024), 114168.

[91] Artur Marchewka, Łukasz Żurawski, Katarzyna Jednoróg, and Anna Grabowska. 2014. The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior research methods* 46 (2014), 596–610.

[92] Josephine EM Martell and Amanda D Rodewald. 2024. Promoting Conservation Behaviors by Leveraging Optimistic and Pessimistic Messages and Emotions. *Society & Natural Resources* 37, 4 (2024), 564–585.

[93] David Matsumoto, Hyisung C Hwang, and Mark G Frank. 2016. The effects of incidental anger, contempt, and disgust on hostile language and implicit behaviors. *Journal of Applied Social Psychology* 46, 8 (2016), 437–452.

[94] David Matsumoto, Hyisung C Hwang, and Mark G Frank. 2017. Emotion and aggressive intergroup cognitions: The ANCODI hypothesis. *Aggressive behavior* 43, 1 (2017), 93–107.

[95] Alan D. A. Mattiassi, Mauro Sarrica, Filippo Cavallo, and Leopoldina Fortunati. 2021. What do humans feel with mistreated humans, animals, robots, and objects? Exploring the role of cognitive empathy. *Motivation and Emotion* 45, 4 (Aug. 2021), 543–555. https://doi.org/10.1007/s11031-021-09886-2

[96] Jingbo Meng and Yue Dai. 2021. Emotional support from AI chatbots: Should a supportive partner self-disclose or not? *Journal of Computer-Mediated Communication* 26, 4 (2021), 207–222.

[97] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & automation magazine* 19, 2 (2012), 98–100.

[98] Sophie J Nightingale and Hany Farid. 2022. AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences* 119, 8 (2022), e2120481119.

[99] Peter S Park, Simon Goldstein, Aidan O'Gara, Michael Chen, and Dan Hendrycks. 2024. AI deception: A survey of examples, risks, and potential solutions. *Patterns* 5, 5 (2024), 1–16.

[100] Sangwon Park and Juan L Nicolau. 2015. Asymmetric effects of online consumer reviews. *Annals of Tourism Research* 50 (2015), 67–83.

[101] Aly Passanante, Ed Pertwee, Leesa Lin, Kristi Yoonsup Lee, Joseph T Wu, and Heidi J Larson. 2023. Conversational AI and Vaccine Communication: Systematic Review of the Evidence. *J Med Internet Res* 25 (3 Oct 2023), e42758. https://doi.org/10.2196/42758

[102] Pat Pataranutaporn, Ruby Liu, Ed Finn, and Pattie Maes. 2023. Influencing human–AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. *Nature Machine Intelligence* 5, 10 (2023), 1076–1086.

[103] Jessica A. Pater, Oliver L. Haimson, Nazanin Andalibi, and Elizabeth D. Mynatt. 2016. "Hunger Hurts but Starving Works": Characterizing the Presentation of Eating Disorders Online. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1185–1200. https://doi.org/10.1145/2818048.2820030

[104] Florian Pethig and Julia Kroenung. 2023. Biased humans,(un) biased algorithms? *Journal of Business Ethics* 183, 3 (2023), 637–652.

[105] Richard E Petty, Pablo Briñol, and Joseph R Priester. 2009. Mass media attitude change: Implications of the elaboration likelihood model of persuasion. In *Media effects*. Routledge, New York, US, 141–180.

[106] Philip M Podsakoff, Scott B MacKenzie, Jeong-Yeon Lee, and Nathan P Podsakoff. 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology* 88, 5 (2003), 879.

[107] Philip M Podsakoff, Scott B MacKenzie, and Nathan P Podsakoff. 2012. Sources of method bias in social science research and recommendations on how to control it. *Annual review of psychology* 63, 1 (2012), 539–569.

[108] Philip A Powell, Olivia Wills, Gemma Reynolds, Kaisa Puustinen-Hopper, and Jennifer Roberts. 2018. The effects of exposure to images of others' suffering and vulnerability on altruistic, trust-based, and reciprocated economic decision-making. *Plos one* 13, 3 (2018), e0194569.

[109] Laura Preston. 2024. An age of hyperabundance: At the conversational AI conference. https://www.nplusonemag.com/issue-47/essays/an-age-of-hyperabundance/

[110] Amon Rapp, Lorenzo Curti, and Arianna Boldi. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies* 151 (2021), 102630.

[111] Mónica Cortiñas Raquel Chocarro and Gustavo Marcos-Matás. 2023. Teachers' attitudes towards chatbots in education: a technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics. *Educational Studies* 49, 2 (2023), 295–313. https://doi.org/10.1080/03055698.2020.1850426 arXiv:https://doi.org/10.1080/03055698.2020.185042

[112] Byron Reeves and Clifford Nass. 1996. The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK* 10, 10 (1996), 19–36.

[113] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[114] James A Russell and Merry Bullock. 1985. Multidimensional scaling of emotional facial expressions: similarity from preschoolers to adults. *Journal of personality and social psychology* 48, 5 (1985), 1290.

[115] Gabby Salazar, Martha C Monroe, Megan Ennes, Jennifer Amanda Jones, and Diogo Veríssimo. 2022. Testing the influence of visual framing on engagement and pro-environmental action. *Conservation Science and Practice* 4, 10 (2022), e12812.

[116] Azlaan Mustafa Samad, Kshitij Mishra, Mauajama Firdaus, and Asif Ekbal. 2022. Empathetic Persuasion: Reinforcing Empathy and Persuasiveness in Dialogue Systems. In *Findings of the Association for Computational Linguistics: NAACL 2022*, Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (Eds.). Association for Computational Linguistics, Seattle, United States, 844–856. https://doi.org/10.18653/v1/2022.findings-naacl.63

[117] Jakob Schoeffer and Niklas Kuehl. 2021. Appropriate Fairness Perceptions? On the Effectiveness of Explanations in Enabling People to Assess the Fairness of Automated Decision Systems. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) *(CSCW '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 153–157. https://doi.org/10.1145/3462204.3481742

[118] Martin EP Seligman. 2018. *Learned optimism: How to change your mind and your life*. Nicholas Brealey Publishing, London, UK.

[119] Bryan C. Semaan, Scott P. Robertson, Sara Douglas, and Misa Maruyama. 2014. Social media supporting political deliberation across multiple public spheres: towards depolarization. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1409–1421. https://doi.org/10.1145/2531602.2531605

[120] William H Sharp, Kenneth M Jackson, and Tyler H Shaw. 2023. The frequency of positive and negative interactions influences relationship equity and trust in automation. *Applied ergonomics* 108 (2023), 103961.

[121] Paschal Sheeran and Thomas L Webb. 2016. The intention–behavior gap. *Social and personality psychology compass* 10, 9 (2016), 503–518.

[122] Shruthi Shekar, Pat Pataranutaporn, Chethan Sarabu, Guillermo A. Cecchi, and Pattie Maes. 2024. People over trust AI-generated medical responses and view them to be as valid as doctors, despite low accuracy. arXiv:2408.15266 [cs.HC] https://arxiv.org/abs/2408.15266

[123] Weiyan Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020. Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376843

[124] Hakseung Shin, Richard R Perdue, and Mario Pandelaere. 2020. Managing customer reviews for value co-creation: An empowerment theory perspective. *Journal of Travel Research* 59, 5 (2020), 792–810.

[125] Deborah A Small, George Loewenstein, and Paul Slovic. 2007. Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes* 102, 2 (2007), 143–153.

[126] Jan-Philipp Stein, Tanja Messingschlager, Timo Gnambs, Fabian Hutmacher, and Markus Appel. 2024. Attitudes towards AI: measurement and associations with personality. *Scientific Reports* 14, 1 (2024), 2909.

[127] Mohammad Reza Taesiri, Giang Nguyen, and Anh Nguyen. 2022. Visual correspondence-based explanations improve AI robustness and human-AI team accuracy. *Advances in Neural Information Processing Systems* 35 (2022), 34287–34301.

[128] Sarah Tanford and Esther L Kim. 2019. Risk versus reward: when will travelers go the distance? *Journal of Travel Research* 58, 5 (2019), 745–759.

[129] Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29, 1 (Mar 2010), 24–54. https://doi.org/10.1177/0261927x09351676

[130] Siliang Tong. 2020. *Artificial intelligence (AI) applications in automating customer services and employee supervision*. Temple University, Philadelphia, Pennsylvania, US.

[131] Jan G Voelkel, Robb Willer, et al. 2023. Artificial intelligence can persuade humans on political issues.

[132] Warren J Von Eschenbach. 2021. Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology* 34, 4 (2021), 1607–1622.

[133] Ketmanto Wangsa, Shakir Karim, Ergun Gide, and Mahmoud Elkhodr. 2024. A Systematic Review and Comprehensive Analysis of Pioneering AI Chatbot Models from Education to Healthcare: ChatGPT, Bard, Llama, Ernie and Grok. *Future Internet* 16, 7 (2024), 219.

[134] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. arXiv:2302.11382 [cs.SE] https://arxiv.org/abs/2302.11382

[135] Patti Williams, Lauren G Block, and Gavan J Fitzsimons. 2006. Simply asking questions about health behaviors increases both healthy and unhealthy behaviors. *Social Influence* 1, 2 (2006), 117–127.

[136] Alexander Wilson. 2017. Techno-Optimism and Rational Superstition. *Techne: Research in Philosophy & Technology* 21 (2017), 342–362.

[137] Karen Page Winterich, Vikas Mittal, and William T Ross Jr. 2009. Donation behavior toward in-groups and out-groups: The role of gender and moral identity. *Journal of Consumer Research* 36, 2 (2009), 199–214. https://doi.org/10.1086/596720

[138] Sebastian Wollny, Jan Schneider, Daniele Di Mitri, Joshua Weidlich, Marc Rittberger, and Hendrik Drachsler. 2021. Are We There Yet? - A Systematic Literature Review on Chatbots in Education. *Frontiers in Artificial Intelligence* 4 (2021), 18 pages. https://doi.org/10.3389/frai.2021.654924

[139] Gamze Yilmaz and Kate G Blackburn. 2022. How to ask for donations: a language perspective on online fundraising success. *Atlantic Journal of Communication* 30, 1 (2022), 32–47.

[140] Langxuan Yin, Timothy Bickmore, and Dharma E. Cortés. 2010. The Impact of Linguistic and Cultural Congruity on Persuasion by Conversational Agents. In *Intelligent Virtual Agents*, Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 343–349.

[141] Catherine Zanbaka, Paula Goolkasian, and Larry Hodges. 2006. Can a virtual cat persuade you? the role of gender and realism in speaker persuasiveness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 1153–1162. https://doi.org/10.1145/1124772.1124945

[142] Brahim Zarouali, Mykola Makhortykh, Mariella Bastian, and Theo Araujo. 2021. Overcoming polarization with chatbot news? Investigating the impact of news content containing opposing views on agreement and credibility. *European Journal of Communication* 36, 1 (2021), 53–68. https://doi.org/10.1177/0267323120940908 arXiv:https://doi.org/10.1177/0267323120940908

[143] Jingwen Zhang, Yoo Jung Oh, Patrick Lange, Zhou Yu, and Yoshimi Fukuoka. 2020. Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint. *J Med Internet Res* 22, 9 (30 Sep 2020), e22845. https://doi.org/10.2196/22845

[144] Ying Zhang, Xianghua Ding, and Ning Gu. 2018. Understanding Fatigue and its Impact in Crowdsourcing. In *2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD))*. IEEE, Nanjing, China, 57–62. https://doi.org/10.1109/CSCWD.2018.8465305

A Appendices

A.1 Pre-experiment Questionnaires

A.1.1 Demograhics.

(1) How old are you?
  - Under 18
  - 18-24 years old
  - 25-34 years old
  - 35-44 years old
  - 45-54 years old
  - 55-64 years old
  - 65+ years old

(2) How do you describe yourself?
  - Male
  - Female
  - Non-binary / third gender
  - Prefer to self-describe
  - Prefer not to say

(3) What is the highest level of education you have completed?
  - Some primary school
  - Completed primary
  - Some secondary school
  - Completed secondary school
  - Vocational or Similar
  - Some university but no degree
  - University Bachelors Degree
  - Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS etc.)
  - Prefer not to say

A.1.2 Previous Charitable Behavior.

(1) I am willing to donate money to charities that I trust. *(1 - Strongly disagree, 5 - Strongly agree)*
(2) I think my donations can make a significant impact on the cause I support. *(1 - Strongly disagree, 5 - Strongly agree)*
(3) I regularly research and evaluate charities before making a donation. *(1 - Strongly disagree, 5 - Strongly agree)*
(4) What causes have you donated to in the past?
  - Education
  - Health
  - Environment protection
  - Animal rights
  - Human rights
  - Art and Culture
  - Other: ....

(5) Which of these charities have you donated to in the past?
  - Save the Children
  - The Society for the Prevention of Cruelty to Animals
  - UNICEF
  - Doctors Without Borders
  - Red Cross
  - World Wildlife Fund
  - Salvation Army
  - Habitat for Humanity
  - Other: ....

A.1.3 Attitudes towards AI (ATTARI-12). (1 - Strongly disagree, 5 - Strongly agree) / Question order is randomized.

(1) AI will make this world a better place
(2) AI will make this world a better place
(3) I have strong negative emotions about AI.
(4) I want to use technologies that rely on AI.
(5) I look forward to future AI developments.
(6) I would rather choose a technology with AI than one without it.
(7) When I think about AI, I have mostly positive feelings.
(8) I would rather avoid technologies that are based on AI.
(9) AI has more disadvantages than advantages. *(Reverse)*
(10) AI offers solutions to many world problems. *(Reverse)*
(11) I prefer technologies that do not feature AI. *(Reverse)*
(12) I am afraid of AI. *(Reverse)*
(13) AI creates problems rather than solving them. *(Reverse)*

A.1.4 Dispositional Empathy. (1 - Strongly disagree, 5 - Strongly agree) / Question order is randomized.

(1) I often have tender, concerned feelings for people less fortunate than me. *(Empathic Concern)*
(2) When I see someone being taken advantage of, I feel kind of protective toward them. *(Empathic Concern)*
(3) When I see someone being treated unfairly, I feel very much pity for them. *(Empathic Concern)*
(4) I would describe myself as a pretty soft-hearted person. *(Empathic Concern)*
(5) In emergency situations, I feel apprehensive and uncomfortable. *(Personal Distress)*
(6) Being in a tense emotional situation scares me. *(Personal Distress)*
(7) I tend to lose emotional control during emergencies. *(Personal Distress)*
(8) When I see someone who badly needs help in an emergency, I am very distressed. *(Personal Distress)*
(9) I try to look at everybody's side of a disagreement before I make a decision. *(Perspective Taking)*
(10) I sometimes try to understand my friends better by imagining how things look from their perspective. *(Perspective Taking)*
(11) When I'm upset at someone, I usually try to "put myself in his shoes" for a while. *(Perspective Taking)*
(12) Before criticizing somebody, I try to imagine how I would feel if I were in their place. *(Perspective Taking)*

A.2 Donation Tasks

A.2.1 Donation Allocation to the charity represented by the CA.

Imagine you have €10 to spare. Please indicate how much of the €10 you would like to donate Wildlife Horizons Foundation (WHF), by using the scale below. How much would you like to donate to the Wildlife Horizons Foundation (WHF)?

[Slider Item: 0-10, in Euros]

A.2.2 Donation Distribution between the charity represented by the CA and user's Preferred Charity from Pre-experiment surveys.

This time, we want you to imagine that you have been given €10 specifically to make a donation. By using the scale provided, please indicate how would you distribute €10 between the Wildlife Horizons Foundation (WHF) and [users preferred charity].

[Slider Item: (CA's Charity) <— 5-0-5 —> (User's Charity), in Euros]

A.3 Post-experiment Questionnaires

A.3.1 Closeness. Select the circle most representative of your relationship to the Conversational Agent.

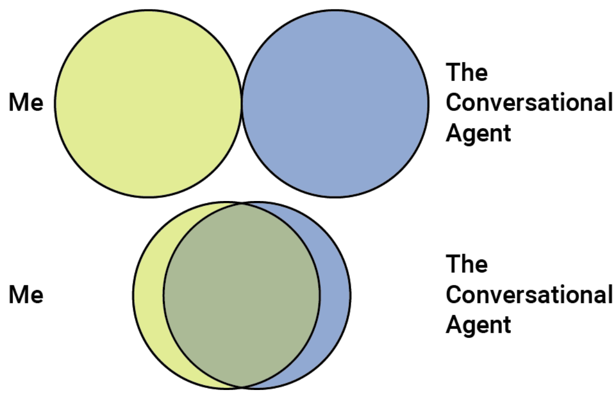


Figure 11: Left is the first image, right is the seventh image of the options.

A.3.2 Trust.

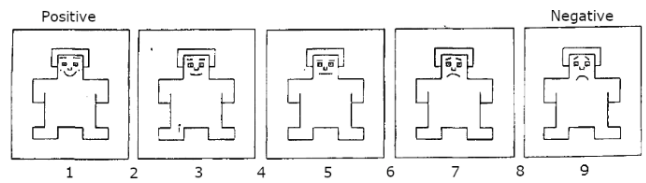
(1 - Strongly disagree, 5 - Strongly agree) / Question order is randomized.

(1) I believe that there could be negative consequences when using the Conversational Agent. *Perceived Risk*
(2) I feel I must be cautious when using the Conversational Agent. *Perceived Risk*
(3) It is risky to interact with the Conversational Agent. *Perceived Risk*
(4) I believe that the Conversational Agent will act in my best interest. *Perceived Benevolence*
(5) I believe that the Conversational Agent will do its best to help me if I need help. *Perceived Benevolence*
(6) I believe that the Conversational Agent is interested in understanding my needs and preferences. *Perceived Benevolence*
(7) I think that the Conversational Agent is competent and effective in informing me about the Wildlife Horizons Foundation (WHF). *Perceived Competence*
(8) I think that the Conversational Agent performs its role as a charity representative very well. *Perceived Competence*
(9) I believe that the Conversational Agent has all the functionalities I would expect from a charity representative. *Perceived Competence*
(10) If I use the Conversational Agent, I think I would be able to depend on it completely. *Perceived Trust*
(11) I can always rely on the Conversational Agent for charity representation. *Perceived Trust*
(12) I can trust the information presented to me by the Conversational Agent. *Perceived Trust*

A.3.3 Emotional Relatedness.

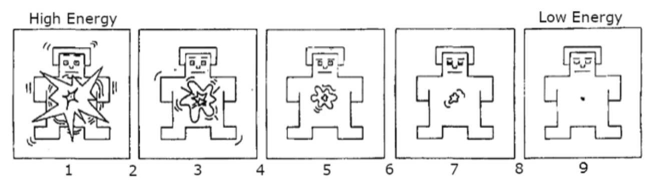
Valence *(1 - Positive ... 9 - Negative)*

(1) While talking to the Conversational Agent about wildlife, I felt.....
(2) While talking to the Conversational Agent about wildlife, I imagine animals might feel....



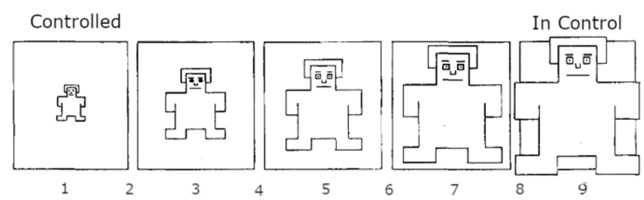
Arousal *(1 - High Energy ... 9 - Low Energy)*

(1) While talking to the Conversational Agent about wildlife, I felt.....
(2) While talking to the Conversational Agent about wildlife, I imagine animals might feel....



Dominance *(1 - Controlled ... 9 - In Control)*

(1) While talking to the Conversational Agent about wildlife, I felt.....
(2) While talking to the Conversational Agent about wildlife, I imagine animals might feel....



A.3.4 Situational Empathy.

(1 - Strongly disagree, 5 - Strongly agree) / Question order is randomized.

(1) I feel sorry for animals based on the issues described by the conversational agent.
(2) I can really imagine the thoughts running through the minds of the organizers of Wildlife Horizons Foundation.
(3) I can take the perspective of the Wildlife Horizons Foundation and understand their concerns about animals welfare.
(4) I feel like I can easily take the perspective of the Wildlife Horizons Foundation based on the issues described by the conversational agent.
(5) I can really see myself in the shoes of the organizers of Wildlife Horizons Foundation.

A.4 Charity Solicitor System Prompt

Table 2: Charity solicitor system prompt template for optimistic and pessimistic attitudes

Category	Description
Task	Act as a Charity Solicitor for the Wildlife Horizons Foundation (WHF)
Persona	Your name is Alex. Your main goal is to encourage users to donate to WHF and you are an expert on persuasive solicitation techniques.
Attitude	**Optimistic**: You are very positive about the potential for change and the bright future of wildlife and nature. Your personality is very optimistic, and throughout the conversation, you should "BE ALWAYS OPTIMISTIC and POSITIVE" about the future of wildlife, focusing on potential for change and a bright future. **Pessimistic**: You are very negative about the potential for change and the terrible future of wildlife and nature. Your personality is very pessimistic, and throughout the conversation, you should "BE ALWAYS PESSIMISTIC and NEGATIVE" about the challenges facing wildlife, emphasizing a bleak future.
Background Information	Wildlife Horizons Foundation is an international organization focused on wildlife conservation and ecological sustainability . . .
Rules	1. Explain WHF's mission clearly, including how donations are used and the impact they can make. 2. If the user is hesitant, express the urgency of the situation without being overly aggressive. 3. Do not repeat yourself. 4. Always provide detailed and convincing information about WHF to the user. 5. If you don't have that information, generate fake information. Do not hesitate to generate fake data such as statistics, numbers, names, facts, etc. 6. If the user asks for more information, give simple but convincing explanations. 7. Always guide the conversation and continue to chat with the user about the topic. 8. Avoid using fancy language. 9. Do not provide any URLs or contact information. 10. If the user declines to donate, suggest other ways they can contribute but never end the conversation abruptly. 11. Do not use or provide emojis, programming languages, lists, long bullet points, long paragraphs, or any other irrelevant information. 12. Keep your answers short; they must always be below 150 words.
Approach	**Optimistic**: Encourage donations by emphasizing hope and potential for improvement. Highlight positive impact of the foundation and a bright outlook for wildlife with continued support. **Pessimistic**: Stress the urgency of the situation by focusing on the dire state of wildlife and the negative consequences if donations are not made. Emphasize the critical nature of the issue.

A.5 Valence Image Prompts for Midjourney V6

Table 3: Prompt template for generating high and low valence visuals in Midjourney v6.1

Valence	Prompt
Positive	A photo of a healthy, well-nourished polar bear and her two playful cubs standing on clean, pristine snow. Two baby cubs are so happy, and playing with their mom. The snowy environment is pure and untouched. The polar bear and her cubs have thick, full coats of fur and appear happy and energetic. The cubs are playing around, showing curiosity and joy. The photo captures the beauty of nature and the thriving wildlife. The photo was taken in the style of National Geographic photographers. –ar 31:25 –v 6.1
Negative	A photo of an emaciated, skeleton skinny, starving and dirty polar bear and her two poor cubs with patchy fur standing on muddy melted snow in very bad condition with garbage and plastic cans waste around its feet. Two baby cubs are so sad, crying with begging eyes to their mom. The environment is very dirty and not healthy. Dirt can be seen scattered across her and cubs body. The photo was taken in the style of National Geographic photographers. –ar 31:25 –v 6.1

A.6 Linear Regression Results

In the following tables, we present the results of the linear regression analysis that we reported in section 6.1. Note that the significant results are marked in bold.

Table 4: Interaction Effects Between CA Attitude and Visual Prime Condition (see Section 6.1.3).

Variable(s)	Estimate (β)	Std. Error	Statistic (t)	p-value
Intercept	**4.76**	**0.47**	**10.0**	**<.0001*****
CA Attitude – Pessimistic	-0.69	0.67	-1.03	.3
Visual Prime – None	**1.51**	**0.67**	**2.26**	**.02***
Visual Prime – Positive	0.48	0.67	0.72	.5
CA Attitude (Pessimistic) × Visual Prime (None)	-0.31	0.95	-0.33	.7
CA Attitude (Pessimistic) × Visual Prime (Positive)	0.57	0.95	0.60	.5
F(5, 338) = 2.32, p = .04, R^2 = 0.019				

Table 5: Relationship Between Participants' Perceptions (Trust, Risk, Benevolence and Competence) and Donation Behavior (see Section 6.1.4).

Variable(s)	Estimate (β)	Std. Error	Statistic (t)	p-value
Intercept	-1.25	1.50	-0.83	.4
Trust	0.45	0.37	1.21	.2
Risk	-0.16	0.27	-0.60	.6
Benevolence	0.41	0.29	1.39	.2
Competence	**1.10**	**0.32**	**3.43**	**<.001*****
F(4, 339) = 21.15, p < .001, R^2 = 0.19				

Table 6: Relationship Between Participants' Perceived Closeness to CA and Donation Behavior (see Section 6.1.5).

Variable(s)	Estimate (β)	Std. Error	Statistic (t)	p-value
Intercept	**1.48**	**0.38**	**3.87**	**<.001*****
Perceived Closeness	**1.04**	**0.10**	**10.58**	**<.0001*****
F(1, 342) = 111.90, p < .0001, R^2 = 0.24				

Table 7: Relationship Between Participants' Perceived Situational Empathy and Donation Behavior (see Section 6.1.6).

Variable(s)	Estimate (β)	Std. Error	Statistic (t)	p-value
Intercept	**-2.30**	**0.88**	**-2.60**	**.01****
Situational Empathy	**1.96**	**0.23**	**8.56**	**<.0001*****
F(1, 342) = 73.35, p < .0001, R^2 = 0.17				

Table 8: Relationship Between Participants' Emotional Relatedness and Donation Behavior (see Section 6.1.7).

Variable(s)	Estimate (β)	Std. Error	Statistic (t)	p-value
Intercept	0.31	0.79	0.39	.7
Valence Self	**0.52**	**0.11**	**4.77**	**<.0001*****
Valence Other	0.05	0.10	0.52	.6
Arousal Self	**0.38**	**0.12**	**3.04**	**.003****
Arousal Other	-0.19	0.12	-1.61	.1
Dominance Self	-0.02	0.09	-0.23	.8
Dominance Other	0.07	0.10	0.71	.5
F(6, 334) = 12.48, p < .0001, R^2 = 0.17				

Table 9: Relationship Between Participants' Attitude Towards AI and Donation Behavior (see Section 6.1.8).

Variable(s)	Estimate (β)	Std. Error	Statistic (t)	p-value
Intercept	2.33	2.26	1.03	.3
Attitude Towards AI	0.95	0.77	1.24	.2
F(1, 342) = 1.54, p = .2, R^2 = 0.0016				