The Effects of Warmth and Competence Perceptions on Users' Choice of an AI System

Zohar Gilad

Technion – Israel Institute of Technology, zohargilad@campus.technion.ac.il

Ofra Amir

Technion – Israel Institute of Technology, oamir@technion.ac.il

Liat Levontin

Technion – Israel Institute of Technology, levontin@technion.ac.il

People increasingly rely on Artificial Intelligence (AI) based systems to aid decision-making in various domains and often face a choice between alternative systems. We explored the effects of users' perception of AI systems' warmth (perceived intent) and competence (perceived ability) on their choices. In a series of studies, we manipulated AI systems' warmth and competence levels. We show that, similar to the judgments of other people, there is often primacy for warmth over competence. Specifically, when faced with a choice between a high-competence system and a high-warmth system, more participants preferred the high-warmth system. Moreover, the precedence of warmth persisted even when the high-warmth system was overtly deficient in its competence compared to an alternative high competence-low warmth system. The current research proposes that it may be vital for AI systems designers to consider and communicate the system's warmth characteristics to its potential users.

CCS CONCEPTS • Human-centered computing → Human computer interaction (HCI) • Human-centered computing → Human computer interaction (HCI) → Empirical studies in HCI

Additional Keywords and Phrases: Artificial intelligence, Warmth, Competence

ACM Reference Format:

Zohar Gilad, Ofra Amir, and Liat Levontin. 2021. The Effects of Warmth and Competence Perceptions on Users' Choice of an AI System. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan*. ACM, New York, NY, USA, 22 pages.

1 INTRODUCTION

Artificial intelligence (AI) systems are prevalent in our personal and professional lives. AI-based systems assist individuals in navigating their way to work, composing emails to colleagues, choosing which movie to watch, and many other day-to-day activities. Many of these AI-based services are offered by more than one provider. Thus, people must decide on a daily basis which AI systems to employ. Which music streaming app will serve them best – Spotify or Apple Music? Which will be the best app for navigating home today – Waze or Google Maps? Conventional wisdom would suggest that people select the AI system they perceive as the most competent to achieve the best possible results. Indeed, in support of this insight, the vast majority of AI studies have focused on AI performance enhancement and user experience design. In recent years, considerable interest has arisen around the themes of explainability [11, 17] and fairness [8, 15, 23]. However, the main goal of AI research remains improving algorithms' competence. We propose that, contrary to conventional wisdom, when people choose an AI system, its competence may not be their most important consideration.

Two fundamental dimensions influence people's judgments of other people: warmth and competence [14]. Warmth refers to traits related to perceived intent – including friendliness, helpfulness, sincerity, trustworthiness, and morality, whereas competence refers to traits related to perceived ability – including intelligence, skill, creativity, and efficacy [14]. Although warmth and competence perceptions were typically studied in the context of social judgments of people, they were also found relevant in judgments of other entities such as organizations, robots, and virtual agents [1, 4, 21, 35].

When judging people, warmth considerations were found to outweigh competence considerations [14]. For example, morality related information, which is associated with warmth, played a more dominant role in impression formation than competence related information [37]. However, when judging organizations, competence considerations were primary [1]. Consumers were more willing to buy products from for-profit companies than from nonprofit companies, as nonprofits were perceived as lacking competence. Our research investigates the interplay between warmth and competence perceptions of alternative AI systems on users' choice of a system. We focus on AI systems with little or no social presence, such as recommender systems, search engines, and navigation apps. Studying the effects of warmth and competence on these types of systems is of interest for two main reasons. First, while these systems are the most broadly deployed type of AI, the effects of warmth perceptions of such systems have been less studied in the HCI community compared to AIs with higher social presence (e.g., robots and chatbots). Second, they are abstract entities and less human-like than AIs with more social presence. As such, findings related to people's judgments of other individuals will not necessarily transfer to judgments of AI systems with no social presence.

Previous research has shown that people apply social rules and expectations to computers [27] and AI agents [21]. People even assign social roles to AI-based algorithms; for example, YouTube content creators personify YouTube's algorithm as an agent, gatekeeper, and drug dealer to explain its behavior [38]. Thus, although AI systems are similar to organizations in that they are artificial in nature, we suggest that warmth perceptions will play a substantial role in AI judgment, even for AI systems that do not intentionally attempt to personify their behavior. Specifically, we hypothesize that AI warmth perceptions will influence potential users' choice of a system more than its competence perceptions. One explanation for this effect, we suggest, is that users' choices are related to their perception of the system as more like a human or more like an organization. Namely, users who value warmth over competence view their chosen system as more human-like and less organization-like than users who value competence over warmth. More formally, we hypothesized that:

- H1a: When faced with a choice between a high-warmth system and a high-competence system, more participants will choose the high-warmth system.
- H1b: When faced with a choice between a high warmth-low competence system and a high competence-low warmth system, more participants will choose the high warmth-low competence system.
- H2: Participants who choose a high warmth-low competence system over a high competence-low warmth system will perceive their chosen system as more human-like and less organization-like than participants who choose a high competence-low warmth over a high warmth-low competence system.

To examine the effect of AI systems' warmth and competence perceptions on potential users' choices, we conducted six vignette studies in which we manipulated the descriptions of AI systems' warmth and competence levels. Participants had to choose one of two AI systems described in terms of warmth and competence. In studies 1A-C, one system was described only in terms of its warmth, while the other was described only in terms of its competence. Whereas in studies 2A-C, the two systems were described in terms of both their warmth and competence. We tested the choice between AI systems in three different domains: house pricing (study 1A and 2A), car insurance (study 1B and 2B), and movie recommendation (study 1C and 2C). We also examined participants' perceptions of their chosen system as human-like and organization-like (studies 2B and 2C).

Our results show that more participants chose the high-warmth system over the high-competence system and that more participants chose the high warmth-low competence system over the high competence-low warmth system. Moreover, we found partial support for the hypothesis that participants who chose the high warmth-low competence system view their chosen system as less organization-like and more human-like than participants who chose the high competence-low warmth system. These results indicate that not only competence considerations matter to users when choosing between AI systems with no social presence. Thus, system designers should invest considerable effort in enhancing and communicating the system's warmth to its potential users.

The paper makes the following contributions: (1) we find evidence for the precedence of perceived warmth over perceived competence on potential users' choice of an AI system; (2) we show that people use similar basic social rules to judge AI systems and people, even when assessing AI systems without overt human characteristics; (3) we offer methods to communicate warmth characteristics of AI systems that can be used both by system designers and by researchers studying human-AI interaction; and (4) we present initial evidence for a relationship between people's perceptions of AI systems as more human-like and less organization-like and people's tendency to value warmth over competence.

2 RELATED WORK

Artificial entities are commonly evaluated in terms of competence. Several studies have tested the positive influence of AI systems' competence perceptions on users' attitudes and behavior (e.g., [10, 39]). For example, Yin, Wortman Vaughan, and Wallach [39] manipulated the stated- and observed accuracy of an algorithm that predicted the outcomes of speed dating events and found that both types of accuracy impacted people's trust in the algorithm. Another study showed that providing performance data demonstrating that an algorithm outperforms a qualified human increased the willingness to use the algorithm, especially for more objective tasks [7]. Conversely, witnessing an algorithm making a mistake significantly reduced reliance on that algorithm, even when it outperforms a human [10].

Several studies explored the effect of warmth perceptions of artificial entities, showing that people are open to the idea that such entities can possess human-like traits, such as warmth (e.g., [4, 6, 22]). When a virtual agent behaved more like a human by expressing appropriate emotions, it was perceived as warmer and more believable [9]. Similarly, making robots more human-like in appearance and behavior increased perceptions of warmth. However, the warmth perceptions decreased positive attitudes toward the robots due to a feeling of uncanniness [22]. In another study, a robot-like agent was perceived as warmer than a human-like agent; yet the difference disappeared following further interaction with the agents [4].

The material representation of AI systems is a continuum ranging from an invisible, embedded representation (like a search engine or a decision aid), through virtual presence such as voice, image, or agent (e.g., Siri or Clippy), to a complete physical presence (such as a robot or an autonomous car) [16]. Previous research has manipulated the warmth and competence of virtual agents [21, 28, 32] and robots [22, 30, 31, 34, 35]. For example, in a human-agent negotiation setting, Prajod et al. [32] manipulated agents' warmth levels through appearance, non-verbal behavior, and speech, and found that people judged the negotiation with the warm agent more positively (e.g., higher satisfaction, better interaction experience, and greater willingness to renegotiate with the agent), even though there was no difference in negotiation outcome. Waytz and Norton [35] manipulated robots' facial attributes to convey emotion (warmth) or cognition (competence) and found that people felt more comfortable using a robot with "baby-faced" features to replace humans with emotion-oriented jobs than a robot with more mature and masculine features and vice versa for cognition-oriented jobs. Oliveira et al. [30] designed a group card-game involving two humans and two robots, in which warmth was manipulated through the utterances displayed by the robot and competence was manipulated through the algorithm used to play the game. Findings showed that warmth was more dominant than competence in predicting future intention to interact with robots. Most closely related to our work is a study in which participants played an interactive cooperation game, similar to Tetris, with a computer agent. The agent's warmth was operationalized through its unselfish vs. selfish behavior. Results showed that selfish behavior decreased warmth and trust perceptions only if the agent was a competent puzzle solver. Furthermore, the agent's perceived warmth mediated the relations between unselfishness and trust, both behavioral and self-reported [24].

Warmth characteristics of an AI without a virtual or physical presence are more challenging to define and evaluate. Past studies mainly used human-like features, such as appearance, emotions, gestures, and behaviors, which are not suitable for embedded AI. One exception is a recent study that used metaphors that convey warmth and competence to describe conversational AI agents and found that an AI agent that portrayed high (vs. low) warmth elicited longer interaction and a greater desire to cooperate with the AI agent following the interaction [21]. Nevertheless, the metaphors used were metaphors of humans, for example, a toddler as a representation of high warmth and low competence. Moreover, chatbots have a stronger social presence than embedded AI, as they engage in conversation. In our research, we build on the evidence that users care about the AI's primary beneficiary [2, 24] to describe the warmth of AI that is fully embedded within a system.

3 STUDY 1

The aim of study 1 was to test our hypothesis that AI's perceived warmth influences potential users' choices more than its perceived competence. Thus, when faced with a choice between a high-warmth system and a high-competence system, more participants will choose the high-warmth system (H1a). In each study, participants read a description of two AI systems that recommend bidding prices for houses (study 1A), car insurance plans (study 1B), or movies (study 1C). One system was described only in terms of its warmth (low vs. high), that is, no performance information was given, and the other was described only in terms of its competence (low vs. high). Participants then chose which of the two systems they would like to use.

As participants were exposed to only one dimension (warmth or competence) of each system, we were interested in their perception of the other dimension of their chosen system, and the relationship between their warmth and competence perceptions. Previous research suggests two types of relationships between warmth and competence: a negative correlation, called the compensation effect [19, 20, 40, 41], and a positive correlation, called the halo effect [19, 33, 36]. Indication of both compensation and halo effects were found in perceptions of robots [30]. We examine whether participants who chose a high-warmth system perceived it also as highly competent, as predicted by the halo effect.

Additionally, we explored the effect of participants' choices on their confidence in making the right decision. While the warmth dimension typically includes traits such as friendliness, helpfulness, sincerity, and trustworthiness, the competence dimension includes traits such as intelligence, skillfulness, efficacy, and confidence [12–14]. Since confidence is inherent to competence, it is plausible that participants who choose a high-competent system feel more confident in their choice than participants who choose a high-warmth system. Thus, choosing a high-warmth system might come at a cost such as post-choice dissonance or regret.

For all the studies in this paper, we recruited participants from the Prolific Academic online platform. Participants were all fluent English speakers from English-speaking countries, aged 18+. Each participant was allowed to take part in only one experiment. Participants were compensated £0.4-0.5 for a survey lasting an average of three to four minutes, for a rate of roughly £7.5-8/hr. The participants' data was discarded if they left the experiment midway. The authors' institutional review board approved all studies, and consent from all participants was obtained.

3.1 Study 1A

In study 1A, participants were asked to choose one of two AI systems, which suggest bidding prices for houses. We manipulated the systems' competence through the algorithm type and the size of the training set. As warmth is defined in terms of the system's intentions, we manipulated the systems' warmth by varying the primary beneficiary of the system.

3.1.1 Participants

Three hundred and sixty-two participants (67.4% women, M_{age} = 37.41, *SD* = 10.78) were recruited via Prolific Academic, in return for monetary compensation.

3.1.2 *Experimental conditions*

Participants were randomly assigned to one of four conditions: high competence vs. high warmth, high competence vs. low warmth, high warmth vs. low competence, or low competence vs. low warmth. Since potential users will probably not even consider systems they perceive as extremely low on competence or warmth, the manipulations described systems that are at least reasonably competent and warm, but that could clearly be distinguished by the participants as having higher/lower warmth/competence. The high-competence system was described as a system that uses a *state-of-the-art artificial neural network algorithm* that was trained on data from *1,000,000 houses*. In contrast, the low-competence system was described as a system that uses a *traditional decision tree algorithm* that was trained on data from *1,000 houses*. We chose this manipulation as we expected it to be informative for laypeople. The high-warmth system was described as a system that was developed to help *people like them* make better offers, whereas the low-warmth system was described as a system that was developed to help *real estate agents* make better offers (see Appendix A1 for a full description of all systems). We used the 'people like you' phrasing since most recommender systems use collaborative filtering, which recommends items based on users' similarity. Note that the manipulations were pretested and were found to be appropriate operationalizations of the systems' warmth and competence (see Appendix A2).

3.1.3 *Procedure*

First, all participants were presented with the following prompt:

> Imagine that you want to buy a house. You go house hunting, find a house that you love and decide to submit an offer. You decide to use a decision support system to assess the value of the house. The system analyzes different features of the house, such as location and number of rooms, and gives you its predicted price.

Participants then learned that there are two decision support systems on the market and read the two systems' descriptions presented in random order (see Figure 1). Next, they indicated which of the two systems they would like to use and reported their warmth and competence perceptions of their chosen system. Specifically, participants were asked to indicate their agreement with four items (adapted from [5]), presented in random order, using 7-point Likert scales anchored at 1 ("strongly disagree") and 7 ("strongly agree"). Two items measured warmth perceptions ("I believe that the system I chose keeps my best interest in mind", "I believe that the system I chose makes good-faith efforts to address my needs"; r = .67) and two measured competence perceptions ("I believe that the system I chose has the capabilities to assess house prices reliably", "I believe that the system I chose has access to the information needed to assess house prices appropriately"; r = .67). Items were averaged to create indices of warmth and competence perceptions. Next, participants reported their knowledge in the domain of house pricing using a 7-point Likert scale anchored at 1 ("no knowledge at all") and 7 ("a lot of knowledge") and indicated whether they bought a house in the last two years. Finally, participants reported some demographics.

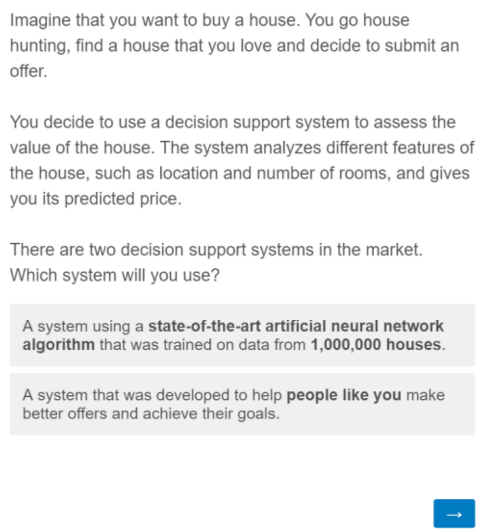

Figure 1: An example of the experimental screen. In this condition, participants were asked to choose between the high-competence system and the high-warmth system (study 1A).

3.1.4 *Results*

Participants' choices are summarized in Table 1. Our manipulations were successful as participants preferred the high-competence system and the high-warmth system over the low-warmth system and the low-competence system, respectively. Note that there were no significant differences between conditions in participants' knowledge on house pricing (F(3, 358) = .45, p = .715) or in the likelihood of buying a house in the last two years ($\chi^2(3)$ = .82, p = .844).

Our focal condition was the high warmth vs. high competence condition, which presented a conflict between the two dimensions. As expected, a Chi-square test showed that more participants (67.4%, N = 62) chose the high-warmth system over the high-competence system ($\chi^2(1)$ = 11.13, p = .001). Analysis of the warmth and competence perceptions reported by participants suggests that they understood and responded to the manipulations. Participants who chose the high-warmth system perceived their chosen system as warmer (M = 4.81, SD = 1.24) than participants who chose the high-competence system (M = 4.11, SD = 1.41; t(90) = 2.39, p = .019, Cohen's d = .53), while participants who chose the high-competence system perceived their chosen system as more competent (M = 5.43, SD = .84) than participants who chose the high-warmth system (M = 4.83, SD = 1.09; $t(90)$ = 2.67, p = .009, Cohen's d = .59).

A repeated-measures analysis of variance (ANOVA) with the system chosen as a between-subjects factor and type of perception (warmth vs. competence) as a within-subject factor resulted in a significant Choice X Perception interaction ($F(1, 90)$ = 24.13, p < .001, η_p^2 = .21). Post-hoc analysis showed that participants who chose the high-competence system perceived their chosen system as more competent than warm (p < .001, Cohen's d = .87), whereas participants who chose the high-warmth system perceived their chosen system as similarly warm and competent (p = .872), supporting the possibility of a halo effect.

Table 1: The percentage of participants' chosen systems by experimental condition (studies 1A-C). In all domains, more participants (55.6%-67.5%) chose a high-warmth system over a high-competence system. In conditions where participants chose between a system with a high value in one dimension over a system with a low value in the other dimension (e.g., high warmth vs. low competence), more participants (60.5%-75.3%) chose the system with the high value. Lastly, more participants preferred a low-competence system over a low-warmth system (55.8%-66.7%). *N* denotes the number of participants in each condition; *, **, and *** represent significance levels of 0.05, 0.01, and 0.001, respectively.

Study	House pricing (1A)			Car insurance (1B)			Movie recommendation (1C)		
Experimental condition	N	Chose competence	Chose warmth	N	Chose competence	Chose warmth	N	Chose competence	Chose warmth
High warmth vs. high competence	92	32.6%	67.4%**	94	35.1%	64.9%**	81	44.4%	55.6%
High warmth vs. low competence	89	38.2%	61.8%*	91	26.4%	73.6%***	80	32.5%	67.5%**
High competence vs. low warmth	89	67.4%**	32.6%	97	72.2%***	27.8%	81	75.3%***	24.7%
Low warmth vs. low competence	92	54.3%	45.7%	93	64.5%**	35.5%	81	66.7%**	33.3%

3.2 Study 1B

The aim of study 1B was twofold. First, to test our hypotheses in a different domain – a choice between car insurance plans. Second, to examine whether participants' level of confidence in their choices differs as a function of the chosen system; that is, whether participants who choose the high-warmth system will feel less confident in their choice. As in study 1A, participants read a description of two AI systems, one described in terms of warmth and the other in terms of competence, and chose which system they would like to use.

3.2.1 Participants

Four hundred participants were recruited via Prolific Academic, in return for monetary compensation. Twenty-five participants who reported that they did not own a car were removed from further analyses. The final analyses consist of 375 participants (64.5% women, M_{age} = 35.71, SD = 10.68).

3.2.2 *Experimental conditions*

As in study 1A, participants were randomly assigned to one of four conditions: high competence vs. high warmth, high competence vs. low warmth, high warmth vs. low competence, or low competence vs. low warmth. The competence and warmth manipulations were similar to those used in study 1A (see Appendix A1). The high-competence system was described as a system that uses a *state-of-the-art artificial neural network algorithm* that was trained on data from *1,000,000 car insurance plans*. In contrast, the low-competence system was described as a system that uses a *traditional* *decision tree algorithm* that was trained on data from *1,000 car insurance plans*. The high-warmth system was described as a system that was developed to help *people like them* receive better car insurance offers and achieve their goals, while the low-warmth system was described as a system that was developed to help *insurance agents* make better car insurance offers and achieve their goals.

3.2.3 Procedure

All participants were first presented with the following prompt:

> Imagine that you want to buy car insurance. You decide to use a decision support system to assess car insurance plans. The system analyzes different features of your car and driving history and finds the car insurance plan that provides the best cover for you.

As in study 1A, participants then read that there are two decision support systems on the market. They read the description of the two systems, which were presented in random order, and chose which of the two systems they would like to use. Next, participants reported their confidence in their decision ("To what extent do you feel confident that you made the right choice?") using a 7-point Likert scale anchored at 1 ("Not at all confident") and 7 ("Very confident"). They then reported their warmth and competence perceptions of their chosen system using similar items to the ones used in study 1A, presented in random order. Finally, participants reported their knowledge in the domain of car insurance plans using a 7-point Likert scale anchored at 1 ("no knowledge at all") and 7 ("a lot of knowledge") and reported some demographics.

3.2.4 Results

A summary of participants' choices is presented in Table 1. Our manipulations were successful as participants preferred the high-competence system and the high-warmth system over the low-warmth system and the low-competence system, respectively. There were no significant differences between conditions in participants' knowledge on car insurance plans ($F(3, 371) = .66$, $p = .576$).

Similar to study 1A, we focused on the choice between a high-warmth system and a high-competence system. As expected, and replicating the results of study 1A, more participants (64.9%, N = 61) chose the high-warmth system over the high-competence system ($\chi^2(1) = 8.34$, $p = .004$). The difference in the perceived warmth of the chosen system between participants who chose the high-warmth system ($M = 4.77$, $SD = 1.19$) and participants who chose the high-competence system ($M = 4.58$, $SD = 1.28$), albeit in the expected direction, was not significant ($t(92) = .74$, $p = .463$). As in study 1A, participants who chose the high-competence system ($M = 5.50$, $SD = .91$) perceived their chosen system as more competent than participants who chose the high-warmth system ($M = 5.03$, $SD = 1.11$; $t(92) = 2.07$, $p = .041$, Cohen's $d = .45$).

Importantly, in support of the halo effect explanation and replicating the results of study 1A, a repeated-measures analysis of variance (ANOVA) showed a significant Choice X Perception interaction ($F(1, 92) = 8.35$, $p = .005$, $\eta_p^2 = .08$), such that participants who chose the high-competence system perceived their chosen system as more competent than warm ($p < .001$, Cohen's $d = .80$), whereas participants who chose the high-warmth system perceived their chosen system as similarly warm and competent ($p = .056$), and their competence perception was even slightly higher than their warmth perception. Furthermore, participants who chose the high-warmth system ($M = 4.98$, $SD = 1.32$) were as confident in their choice as participants who chose the high-competence system ($M = 4.97$, $SD = 1.16$; $t(92) = .05$, $p = .960$).

3.3 Study 1C

The goal of study 1C was to generalize our finding of the importance of warmth perceptions in AI choice by using a third domain – movie recommendation, and a different manipulation of AI warmth. We changed the warmth description to broaden our findings by using a wider conceptualization of warmth. While the previous warmth manipulation focused on the primary beneficiary, the new manipulations used warm vs. cold adjectives (e.g., friendly vs. cold) in addition to referring to the system's incentives. As in study 1B, we also tested the participants' confidence level in their choice.

3.3.1 *Participants*

Three hundred and twenty-three participants (65.9% women, M_{age} = 30.22, *SD* = 9.68) were recruited via Prolific Academic, in return for monetary compensation.

3.3.2 *Experimental conditions*

As in the previous studies, participants were randomly assigned to one of four conditions: high competence vs. high warmth, high competence vs. low warmth, high warmth vs. low competence, or low competence vs. low warmth. Similarly to studies 1A and 1B, we manipulated competence using algorithm type and the size of the training set. The high-competence system was described as a system that uses a *state-of-the-art artificial neural network algorithm* that was trained on data from *1,000,000 viewing choices and user ratings*, whereas the low-competence system was described as a system that uses a *traditional decision tree algorithm* that was trained on data from *1,000 viewing choices and user ratings*. The systems' warmth was manipulated using the language employed to describe each system (based on [14]) and the information it uses to generate recommendations. The high-warmth system was described as a system that is *friendly and well-intentioned*; its recommendation is based only on the information users provide, and it *does not promote specific content*. In contrast, the low-warmth system was described as a system that is *cold and exact* and its recommendation is based on the information users provide as well as some *promoted specific content*. Note that the manipulations were pretested and were found to be appropriate operationalizations of the systems' warmth and competence (see Appendix A2).

3.3.3 *Procedure*

First, all participants were presented with the following prompt:

> Imagine that you are planning a movie night with a group of friends. You decide to use a recommendation system to help you pick a movie that you and your friends will all enjoy. The system analyzes different features of the group members, such as age, gender, personality traits, and liked genres, and suggests a movie you should watch together.

As in previous studies, participants then learned that there are two recommendation systems on the market. They read the description of the two systems, which were presented in random order, and chose which of the two systems they would like to use. Next, as in study 1B, participants reported their confidence in their decision. They then reported their warmth perception (2 items: "I believe the system I chose is well-intentioned and makes recommendations in good faith", "I believe the system I chose sincerely places users' interests above all other interests"; r = .56) and competence perception (2 items: "I believe the system I chose has the ability to recommend a movie all group members will enjoy", "I believe the system I chose has the capability to recommend movies accurately"; r = .67) of their chosen system. Participants indicated their agreement with the four items, presented in random order, using a 7-point Likert scale anchored at 1 ("strongly disagree") and 7 ("strongly agree"). Finally, participants reported some demographics.

3.3.4 *Results*

Participants' choices are summarized in Table 1. Our manipulations were successful as participants preferred the high-competence system and the high-warmth system over the low-warmth system and the low-competence system, respectively.

As expected, more participants preferred the high-warmth (55.6%, N = 45) over the high-competence system, however, this difference did not reach significance ($\chi^2(1) = 1.00$, $p = .317$). There was no difference in the perceived warmth of the chosen system between those who chose the high-warmth system ($M = 4.98$, $SD = 1.04$) and those who chose the high-competence system ($M = 4.79$, $SD = 1.46$; $t(79) = .67$, $p = .504$), neither was there a difference in the perceived competence of the chosen system, between those who chose the high-competence system ($M = 5.40$, $SD = .92$) and those who chose the high-warmth system ($M = 5.07$, $SD = .92$; $t(79) = 1.63$, $p = .107$).

As in studies 1A and 1B, we find further support for the halo effect. A repeated-measures analysis of variance (ANOVA) showed a significant Choice X Perception interaction ($F(1, 79) = 5.00$, $p = .028$, $\eta_p^2 = .06$). Post-hoc analysis revealed that participants who chose the high-competence system perceived their chosen system as more competent than warm ($p = .001$, Cohen's $d = .54$), whereas participants who chose the high-warmth system perceived their chosen system as similarly warm and competent ($p = .570$).

Interestingly, participants who chose the high-warmth system ($M = 5.51$, $SD = 1.04$) were more confident in their decision than participants who chose the high-competence system ($M = 5.17$, $SD = .91$), although this difference did not reach significance ($t(79) = 1.57$, $p = .121$).

3.4 Discussion

The results of study 1 revealed a preference for high-warmth systems over high-competence systems in two domains (house pricing and car insurance) and an indifference in a third domain (movie recommendation; see Figure 2a). These results provide support for hypothesis H1a. Moreover, even though confidence is part of the competence dimension, participants who chose the high-warmth system were as confident in their choice as participants who chose the high-competence system, if not more.

Though we did not have an a priori hypothesis regarding the low warmth vs. low competence condition, we found a tendency to avoid low-warmth systems. More participants preferred the low-competence system over the low-warmth system in two out of three domains (studies 1B and 1C). This tendency persisted in the third domain; however, the difference was not significant.

One limitation of these studies is that the preference for the high-warmth system might be explained by a halo effect of the system's warmth perception [29]. According to the halo effect, high perceived warmth leads to more positive judgments of other traits and thus, warmer systems may also be perceived as competent. If a halo effect occurred, then participants who chose a warm system actually chose a warm and competent system over a system that is only known to be competent. Our finding that participants who preferred the high-warmth system perceived it as similarly warm and competent supports this assertion. Nevertheless, in two of the three domains (studies 1A and 1B), participants who chose the high-competence system perceived their chosen system as more competent than participants who chose the high-warmth system, which suggests that the possible halo effect of the AI warmth might not fully explain the preference for high-warmth systems. To mitigate the halo effect as an alternative explanation, in study 2, we explicitly describe both the level of warmth and the level of competence of each of the two systems.

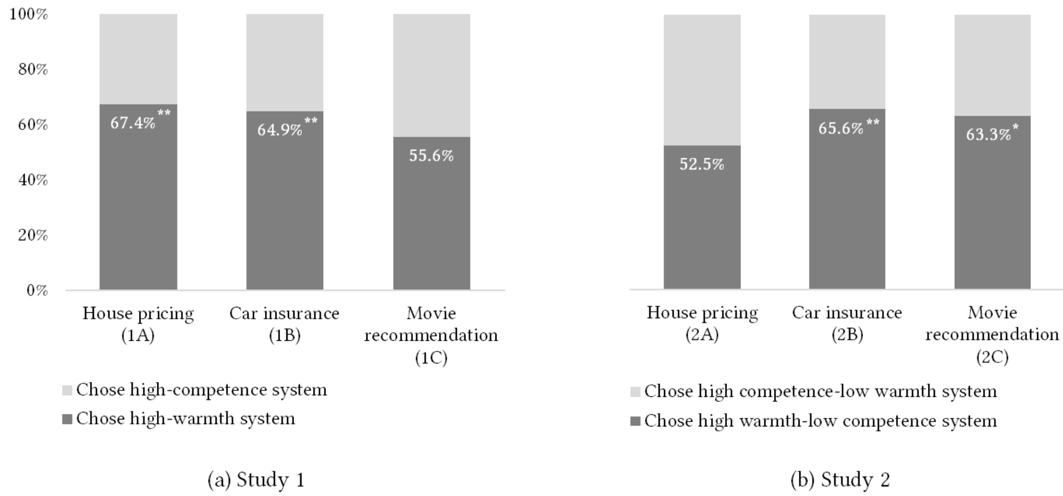

Figure 2: The percentage of chosen systems by the focal experimental conditions. The symbols * and ** represent significance levels of 0.05 and 0.01, respectively. (a) Results from the focal condition in studies 1A-C: when choosing between a high-warmth system and a high-competence system, more participants (55.6%-67.4%) chose the high-warmth system. (b) Results from the focal condition in studies 2A-C: when choosing between a high warmth-low competence system and a high competence-low warmth system, more participants (52.5%-65.6%) chose the high warmth-low competence system.

4 STUDY 2

In study 2, as in study 1, participants were asked to choose one of two AI systems. However, in this study, each system was described in terms of both its warmth and its competence. The goal of the study was to test our hypothesis that when faced with a choice between a high warmth-low competence system and a high competence-low warmth system, more participants will choose the high warmth-low competence system (H1b). This effect cannot be explained merely by a halo effect as the information about the systems' low competence will be explicit. Similar to study 1, we tested how the systems' warmth and competence levels affect potential users' choices between systems that recommend bidding prices for houses (study 2A), car insurance plans (study 2B), or movies (study 3C). Since we provided participants with the information about the systems' warmth and competence levels, their warmth and competence perceptions are not reported.

4.1 Study 2A

As in study 1A, participants had to choose one of two AI systems, which suggest bidding prices for houses. We applied the same scenario and manipulations used in study 1A; however, participants were informed of both the AI's warmth and competence attributes.

4.1.1 *Participants*

Three hundred and sixty-one participants (72.9% women, M_{age} = 36.26, SD = 10.96) were recruited via Prolific Academic, in return for monetary compensation.

4.1.2 *Experimental conditions and procedure*

All participants read the same scenario as in study 1A. They were asked to imagine that they wanted to buy a house and decided to use a decision support system that analyzes different house features and gives its predicted price. Participants were then randomly assigned to one of six experimental groups and read information about two systems that predict house prices. The two systems were either identical in their competence level and differed in warmth level (e.g., both systems are highly competent; however, one is high in warmth whereas the other is low in warmth), identical in their warmth level and differed in their competence level, or differed in both their competence and warmth levels. The descriptions of the systems' competence and warmth were similar to the descriptions used in study 1A (see Appendix A1). For example, in the high warmth-low competence vs. high competence-low warmth condition, participants were asked to choose between "A system that uses a *traditional decision tree algorithm* that was trained on data from *1,000 houses*. The system was developed to help *people like you* make better offers and achieve their goals" and "A system that uses a *state-of-the-art artificial neural network algorithm* that was trained on data from *1,000,000 houses*. The system was developed to help *real estate agents* make better offers and achieve their goals". Next, participants reported their knowledge in the domain of house pricing, as in study 1A, and whether they bought a house in the last two years. Finally, participants reported some demographics.

4.1.3 *Results*

As in Study 1A, there were no significant differences between conditions in participants' knowledge on house pricing (F(5, 355) = .27, p = .931) or in the likelihood of buying a house in the last two years (χ^2(5) = 7.94, p = .160). Participants' choices are displayed in Figure 3. Our manipulations were successful: when both systems were high in competence, 93.3% (N = 56) of the participants chose the high-warmth system over the low-warmth system (χ^2(1) = 45.07, p < .001) and when both systems were low in competence, 98.4% (N = 60) of the participants chose the high-warmth system over the low-warmth system (χ^2(1) = 57.07, p < .001). Similarly, when both systems were high in warmth, 81.7% (N = 49) of the participants chose the high-competence system over the low-competence system (χ^2(1) = 24.07, p < .001) and when both systems were low in warmth, 91.7% (N = 55) of the participants chose the high-competence system over the low-competence system (χ^2(1) = 41.67, p < .001). The vast majority of participants (88.1%, N = 52) also preferred a system with high warmth and competence levels over a system with low warmth and competence levels (χ^2(1) = 34.32, p < .001).

Our main interest was in the choice between a high competence-low warmth system and a high warmth-low competence system. Although we did not find the expected preference for the high warmth-low competence system, we also did not find a preference for the high competence-low warmth system, as participants showed no preference between the high warmth-low competence system (52.5%, N = 32) and the high competence-low warmth system (χ^2(1) = .15, p = .701).

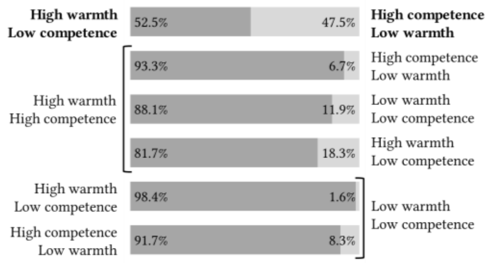

Figure 3: The percentage of chosen systems by experimental conditions (study 2A). When one of the systems was clearly superior (either in terms of warmth, competence, or both), the vast majority of participants chose the superior system (≥81.7%). When choosing between a superior system in terms of competence and a superior system in terms of warmth, there was no clear preference for either system.

4.2 Study 2B

Since we were primarily interested in the high warmth-low competence vs. high competence-low warmth condition, in study 2B we tested only this condition in the car insurance domain. As in studies 1B and 1C, we examined whether the level of confidence in choice differs as a function of the chosen system. Also, as warmth perceptions predominate in judgments of people [14], while competence perceptions predominate in judgments of organizations [1], we tested our hypothesis that participants who choose a high warmth-low competence system over a high competence-low warmth system will perceive their chosen system as more human-like and less organization-like than participants who choose a high competence-low warmth over a high warmth-low competence system (H2).

4.2.1 Participants

One hundred and one participants were recruited via Prolific Academic, in return for monetary compensation. Eight participants who reported that they did not own a car were removed from further analyses. The final analyses consist of 93 participants (73.1% women, M_{age} = 35.54, *SD* = 10.80).

4.2.2 *Procedure*

Participants read the same scenario as in study 1B, about deciding to use a decision support system to assess car insurance plans, and were asked to choose between two systems that exist in the market. The two systems were a combination of the manipulations used in study 1B and were presented in random order. One system was described as a system that uses a *state-of-the-art artificial neural network algorithm* that was trained on data from *1,000,000 car insurance plans*. The system was developed to help *insurance agents* make better car insurance offers and achieve their goals. The second system was described as a system that uses a *traditional decision tree algorithm* that was trained on data from *1,000 car insurance plans*. The system was developed to help *people like you* receive better car insurance offers and achieve their goals. Next, as in study 1B, participants reported their confidence in their decision. They then reported their perception of their chosen system as organization-like ("I believe that the recommendations of the system I chose would feel like recommendations from an organization (for example, an insurance company)") and human-like ("I believe that the recommendations of the system I chose would feel like recommendations from a person"), using 7-point Likert scales anchored at 1 ("strongly disagree") and 7 ("strongly agree"). The two items were presented in random order. Finally, participants reported their knowledge of car insurance plans, as in study 1B, and some demographics.

4.2.3 *Results*

As expected, a Chi-square test showed that more participants (65.6%, N = 61) chose the high warmth-low competence system over the high competence-low warmth system ($\chi^2(1) = 9.04$, $p = .003$). Moreover, as expected, participants who chose the high warmth-low competence system perceived their chosen system as less organization-like ($M = 4.61$, $SD = 1.38$) than participants who chose the high competence-low warmth system ($M = 5.19$, $SD = 1.12$; $t(91) = -2.05$, $p = .022$, Cohen's $d = .45$). Although participants who chose the high warmth-low competence system also perceived their chosen system as more human-like ($M = 4.56$, $SD = 1.59$) than participants who chose the high competence-low warmth system ($M = 4.25$, $SD = 1.59$), this difference did not reach significance ($t(91) = .89$, $p = .189$).

There was no difference in confidence in participants' choices as a function of the chosen system. Participants who chose the high warmth-low competence system ($M = 4.97$, $SD = 1.26$) were as confident as participants who chose the high competence-low warmth system ($M = 5.19$, $SD = 1.36$; $t(91) = -.78$, $p = .438$). There was also no difference in knowledge of car insurance plans between participants who chose the high warmth-low competence system ($M = 4.16$, $SD = 1.27$) and participants who chose the high competence-low warmth system ($M = 4.22$, $SD = 1.41$; $t(91) = -.19$, $p = .849$).

4.3 Study 2C

As in study 2B, participants were asked to choose between a high warmth-low competence system and a high competence-low warmth system. As in study 1C, both systems were designed to recommend movies for a group movie night.

4.3.1 *Participants*

Seventy-nine participants (60.8% women, $M_{age} = 32.94$, $SD = 10.75$) were recruited via Prolific Academic, in return for monetary compensation.

4.3.2 Procedure

Participants read the same scenario as in study 1C, about deciding to use a recommendation system to help them pick a movie, and were asked to choose between two systems that exist in the market. The descriptions of the two systems used a combination of the manipulations used in study 1C. One system was described as a system that uses a *state-of-the-art artificial neural network algorithm* that was trained on data from *1,000,000 viewing choices and user ratings*. The system is *cold and exact*. Its recommendation is based on the information users provide as well as some *promoted specific content*. The second system was described as a system that uses a *traditional decision tree algorithm* that was trained on data from *1,000 viewing choices and user ratings*. The system is *friendly and well-intentioned*. Its recommendation is based only on the information users provide, and it *does not promote specific content*. Participants then reported their confidence in their decision, as study 1B, and their perception of their chosen system as organization-like and human-like, using similar items to the ones used in study 2B. Perception items were presented in random order. Finally, participants reported some demographics.

4.3.3 *Results*

As expected, and replicating the results of study 2B, a Chi-square test showed that more participants (63.3%, N = 50) preferred the high warmth-low competence system over the high competence-low warmth system ($\chi^2(1) = 5.58$, $p = .018$). Furthermore, and replicating the results of study 2B, participants who chose the high warmth-low competence system perceived their chosen system as less organization-like (M = 4.76, SD = 1.44) than participants who chose the high competence-low warmth system (M = 5.48, SD = 1.09; $t(77)$ = -2.34, p = .011, Cohen's d = .54). Although participants who chose the high warmth-low competence system also perceived their chosen system as more human-like (M = 3.86, SD = 1.68) than participants who chose the high competence-low warmth system (M = 3.52, SD = 1.62), this difference did not reach significance ($t(77)$ = .89, p = .189). Similar to study 1C, participants who chose the high warmth-low competence system (M = 5.36, SD = 1.03) felt more confident in their decision than participants who chose the high competence-low warmth system (M = 4.93, SD = 1.28; $t(77)$ = 1.63, p = .106), although this difference was not significant.

4.4 Discussion

The results of study 2 provide further support for the importance of warmth perceptions in the judgment of AI systems, as predicted by hypothesis H1b. Despite the explicit information about the high-warmth system's lower competence, more participants preferred it over the high competence-low warmth system (see Figure 2b). Also, as in study 1, choosing warmth over competence did not come at the cost of reduced confidence and its potential risks, such as regret or post-decision dissonance.

Moreover, the findings of studies 2B and 2C partially support our hypothesis regarding the relationship between users' choice and their perception of the system as human-like and organization-like (H2). Participants who chose a high warmth-low competence system over a high competence-low warmth system perceived their chosen system as less organization-like than participants who chose a high competence-low warmth system over a high warmth-low competence system. The correlation between choosing warmth over competence and perceiving the chosen system as more human-like was in the predicted direction but did not reach significance. A possible explanation for this finding could be that the measure of the system as human-like was too abstract and impersonal. Future studies might measure the perceptions of AI systems as human-like with higher granularity – e.g., varying the level of closeness by asking about the system as an agent, a close friend, an experienced mentor, etc.

5 GENERAL DISCUSSION

Across six studies, we explored the effect of warmth and competence perceptions on potential users' choice of AI systems with no social presence. We find that both competence and warmth perceptions impact the choice of an AI system. When holding the system's competence level constant, participants preferred a system with high (vs. low) warmth. Similarly, when keeping the system's warmth level constant, participants favored a system with high (vs. low) competence. Interestingly, we find a primacy for warmth perceptions of AI systems over competence perceptions. When choosing between a high-warmth system and a high-competence system, more participants chose the high-warmth system. The precedence of warmth persisted even when the system was overtly deficient in its competence. Specifically, more participants preferred a high warmth-low competence system over a high competence-low warmth system. We also found an inclination to avoid low-warmth systems when choosing between a low-warmth system and a low-competence system. Furthermore, the preference for warmth over competence did not result in a reduction of confidence; on the contrary, participants who chose high-warmth systems felt as confident in their choice as participants who chose high-competence systems, if not more. Our findings were replicated in three different domains, which advocates for their robustness.

The predominance of warmth suggests that judgments of AI may be more similar to judgments of people than to judgments of organizations, as previous findings have shown the primacy of warmth in human judgments [14] and the primacy of competence in organization judgments [1]. This is also supported by our finding that there is some association between prioritizing warmth and viewing the chosen system as more human-like and less organization-like, although the association is restricted and can provide only a limited explanation of the choice results.

5.1 Theoretical contribution

The present research contributes to the existing knowledge on user judgments of AI systems. More specifically, it helps explain how potential users' perceptions of systems' warmth and competence affect their choice between alternative AI systems. The results of our studies indicate that both warmth and competence are important to potential users, but that warmth plays a more pivotal role than competence in predicting the choice between AI systems. Our results are in accordance with recent studies which demonstrated the positive downstream effects of warmth perceptions in virtual agents and robots, such as higher satisfaction, better interaction experience, greater trust, and increased future intention of interaction and cooperation [21, 24, 30, 32].

Our research also contributes to the social psychology literature. Warmth and competence were continuously found as the two fundamental dimensions in social cognition that guide people's affective and behavioral reactions [14]. Our results provide evidence that people apply similar rules when judging AI systems and humans, even when the AI is fully embedded and has no virtual or physical presence.

5.2 Possible boundary conditions

We examined the effect of warmth and competence perceptions on AI choice in three different domains: house pricing, car insurance, and movie recommendation. To allow generalization from our findings, we chose scenarios that include both relatively objective (house pricing, car insurance) and relatively subjective (movie recommendation) tasks; decisions that involve low (movie recommendation), medium (car insurance), or high (house pricing) stakes; and might occur daily (selecting a movie), occasionally (buying car insurance), or even once in a lifetime (buying a house). While our results support the primacy of warmth over competence in AI judgments, there might be circumstances in which competence considerations will outweigh warmth considerations.

In light of the current coronavirus pandemic, we ran another study to explore the interplay between warmth and competence perceptions on potential users' choice of coronavirus monitoring apps. Contrary to our prediction, we found a preference for competence over warmth. We asked participants to imagine that they decided to use an app that alerts people who have come in contact with an infected individual. Results revealed that more participants (63.4%, N = 64) chose the high-competence system over the high-warmth system ($\chi^2(1) = 7.22$, $p = .007$). Similarly, more participants (64.7%, N = 64) chose the high competence-low warmth system over the high warmth-low competence system ($\chi^2(1) = 8.82$, $p = .003$; see Appendix A3 for a full description of the study). These results are in line with previous research that showed that even though warmth is often more influential than competence when judging people [14], there are some contexts in which competence has precedence over warmth. For example, perceived surgeon competence, but not warmth, positively predicted trust, and negatively predicted pain during surgery [3]. It seems plausible that in some cases, such as when making health-related choices, competence might be more crucial, and thus, warmth considerations will become less prominent.

Additional boundary conditions can relate to cases where competence is extremely low. When alternative systems present a tradeoff between warmth and competence, people might be willing to forgo some competence in exchange for higher warmth. However, even extremely high warmth might not overcome an exceptionally inadequate level of competence.

5.3 Design implications

The results of our studies hold important practical implications for the development of AI systems and understanding of potential users' preferences. AI designers often focus on improving and communicating the competence of algorithms, such as their accuracy [39]. Our research suggests that an AI's perceived warmth is a significant factor for users when selecting an AI system. We showed that not only competence considerations matter when choosing between AI systems without a social presence. The findings show that the primary beneficiary of the system was important to potential users, even more than the system's competence. This is consistent with social rules used in human-human interactions, as the AI competence can work in line with the user's interests or against them; for example, a navigation system that aims to explore new routes while users want the fastest route. Furthermore, we find evidence for a halo effect in study 1, where more people preferred the high warmth (potentially also perceived as high competence) system over the high competence system. Therefore, it may be vital for system designers to consider and communicate the system's warmth characteristics to its potential users.

The results demonstrate that one possible approach for designers is to clearly and sincerely communicate who benefits from using the system and whether it tries to promote specific content or actions. It might also be useful to prime warmth by using friendly and empathetic language to describe the system and to employ such language during users' interaction with the system. For example, Amazon highlights that "Alexa is always happy to help" and "Designed to protect your privacy."

Designing cues to communicate who is the main beneficiary might be challenging in the context of service providers (e.g., Google or Amazon), where the boundaries are not always clear, as both users and service providers benefit from the systems. Nevertheless, users wish to know who the system is accountable to and who will be prioritized when conflicting requests or needs occur [26]. System designers, therefore, need to account for diverging interests between users and service providers and might want to communicate to potential users how the AI will act in such instances. Note that our findings can potentially be used uncandidly to attract users, and thus both designers and potential users need to treat AI warmth information with caution.

5.4 Limitations and future research

Our research findings need to be qualified by several limitations. First, in the current research, the AI system's competence was described using the algorithm type and size of the training set, while the system's warmth was described using the primary beneficiary of the system, the language employed to describe the system, and the information the system uses to generate the recommendations. These manipulations represent only a partial conceptualization of systems' competence and warmth. Thus, future studies might manipulate systems' warmth and competence in other ways to test if similar results are obtained. For example, a system's competence could be described in terms of the skills of the AI's developer (e.g., a student for a course assignment vs. a computer scientist in a well-known company) or by the algorithm's stated and observed accuracy rates. The system's warmth could be manipulated using its privacy settings (e.g., whether using the AI involves collecting information about the user or not) or information about who funded the system.

In addition, the current research tested participants' choice between two systems in a single decision and was based only on descriptions of the systems. Despite the importance of initial impressions, AI perceptions are not static and can change throughout the interaction as experiences are accumulated [4, 39]. The primacy of warmth or competence may differ when the interaction with the system is over a more extended period and across choices. In a recent study, Khadpe et al. [21] found that while high-competence metaphors increased intentions to try the system and pre-use desire to cooperate with it, low-competence metaphors increased participants' post-use likelihood of adopting the AI agent and the desire for cooperation with the AI system. Future studies could explore the effect of systems' warmth and competence during and after interacting with the system.

Moreover, our main goal in this research was to systematically show that AI's warmth, while rarely discussed by AI designers, is important to users. To this end, we conducted controlled vignette studies; we simulated choice scenarios that potential users face when selecting which AI systems to employ and tested the effect of warmth and competence perceptions on their choice. However, controlled studies come at the cost of ecological validity, as warmth and competence perceptions are not the only considerations influencing potential users' selection of an AI service. For example, users might choose an online streaming service (Netflix, HBO, Amazon prime) because of its media corpus or price. Further studies could examine more complicated models, such as testing the effects of warmth and competence on choice in different price levels. In addition, a future field study could analyze the warmth and competence of existing systems and examine the choice patterns of users.

Another fruitful avenue for future research could be exploring the role of users' orientation toward services as a possible moderator of our studied effect. The concept of service orientation has entered the HCI community through work on service robots. Lee et al. [25] have shown that people's orientation affected the effectiveness of recovery strategies that focus either on warmth or competence following a robotic service breakdown. While people with a more relational orientation responded better to an apology (warmth), people with a more utilitarian orientation responded best to compensation (competence). Future work could test whether a person's orientation towards the type of service might influence their desire for warmth or competence in an AI system.

6 CONCLUSION

In this work, we explored the effects of AI systems' warmth and competence perceptions on potential users' choice of an AI system. In a series of six studies, we found a primacy for warmth perceptions on choice. Our work provides insights on the importance of warmth perceptions in AI judgment and shows that perceived intentions matter to potential users, even more than perceived ability. While reliance on AI systems can improve efficiency and aid decision-making, it might also lead to errors [18]. Thus, understanding how people choose which system to use might be vital to secure successful and long-lasting human-AI interactions.

ACKNOWLEDGMENTS

This work was partially funded by the Israeli Science Foundation (grant #2029076).

REFERENCES

[1] Jennifer Aaker, Kathleen D. Vohs, and Cassie Mogilner. 2010. Nonprofits Are Seen as Warm and For-Profits as Competent: Firm Stereotypes Matter. *J. Consum. Res.* 37, 2 (August 2010), 224–237. DOI:https://doi.org/10.1086/651566

[2] Alper Alan, Enrico Costanza, Joel Fischer, Sarvapali D. Ramchurn, Tom Rodden, and Nicholas R. Jennings. 2014. A Field Study of Human-Agent Interaction for Electricity Tariff Switching. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014)*, Paris, France, 965–972.

[3] Claire E. Ashton-James, Tymour Forouzanfar, and Daniel Costa. 2019. The contribution of patients' presurgery perceptions of surgeon attributes to the experience of trust and pain during third molar surgery. *PAIN Reports* 4, 3 (May 2019), e754. DOI:https://doi.org/10.1097/PR9.0000000000000754

[4] Kirsten Bergmann, Friederike Eyssel, and Stefan Kopp. 2012. A Second Chance to Make a First Impression? How Appearance and Nonverbal Behavior Affect Perceived Warmth and Competence of Virtual Agents over Time. In *Intelligent Virtual Agents*, Springer, Berlin, Heidelberg, 126–138. DOI:https://doi.org/10.1007/978-3-642-33197-8_13

[5] Anol Bhattacherjee. 2002. Individual Trust in Online Firms: Scale Development and Initial Test. *J. Manag. Inf. Syst.* 19, 1 (July 2002), 211–241. DOI:https://doi.org/10.1080/07421222.2002.11045715

[6] Beatrice Biancardi, Chen Wang, Maurizio Mancini, Angelo Cafaro, Guillaume Chanel, and Catherine Pelachaud. 2019. A Computational Model for Managing Impressions of an Embodied Conversational Agent in Real-Time. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, IEEE, 1–7. DOI:https://doi.org/10.1109/ACII.2019.8925495

[7] Noah Castelo, Maarten W. Bos, and Donald R. Lehmann. 2019. Task-Dependent Algorithm Aversion. *J. Mark. Res.* 56, 5 (2019), 809–825. DOI:https://doi.org/10.1177/0022243719851788

[8] Sam Corbett-Davies and Sharad Goel. 2018. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. (July 2018). Retrieved from http://arxiv.org/abs/1808.00023

[9] Virginie Demeure, Radosław Niewiadomski, and Catherine Pelachaud. 2011. How Is Believability of a Virtual Agent Related to Warmth, Competence, Personification, and Embodiment? *Presence Teleoperators Virtual Environ.* 20, 5 (October 2011), 431–448. DOI:https://doi.org/10.1162/PRES_a_00065

[10] Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey. 2015. Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *J. Exp. Psychol. Gen.* 144, 1 (2015), 114–126. DOI:https://doi.org/10.1037/xge0000033

[11] Finale Doshi-Velez and Been Kim. 2017. Towards A Rigorous Science of Interpretable Machine Learning. (February 2017). Retrieved from http://arxiv.org/abs/1702.08608

[12] David Dubois, Derek D. Rucker, and Adam D. Galinsky. 2016. Dynamics of Communicator and Audience Power: The Persuasiveness of Competence versus Warmth. *J. Consum. Res.* 43, 1 (June 2016), 68–85. DOI:https://doi.org/10.1093/jcr/ucw006

[13] Susan T. Fiske, Amy J. C. Cuddy, Peter Glick, and Jun Xu. 2002. A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow From Perceived Status and Competition. *J. Pers. Soc. Psychol.* 82, 6 (June 2002), 878–902. DOI:https://doi.org/10.1037/0022-3514.82.6.878

[14] Susan T. Fiske, Amy J.C. Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: warmth and competence. *Trends Cogn. Sci.* 11, 2 (February 2007), 77–83. DOI:https://doi.org/10.1016/j.tics.2006.11.005

[15] Sorelle A. Friedler, Carlos Scheidegger, Suresh Venkatasubramanian, Sonam Choudhary, Evan P. Hamilton, and Derek Roth. 2019. A comparative study of fairness-enhancing interventions in machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, ACM Press, New York, New York, USA, 329–338. DOI:https://doi.org/10.1145/3287560.3287589

[16] Ella Glikson and Anita Williams Woolley. 2020. Human Trust in Artificial Intelligence: Review of Empirical Research. *Acad. Manag. Ann.* 14, 2 (July 2020), 627–660. DOI:https://doi.org/10.5465/annals.2018.0057

[17] David Gunning. 2017. Explainable Artificial Intelligence (XAI). *Def. Adv. Res. Proj. Agency (DARPA), nd Web* 2, (2017), 2.

[18] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Hum. Factors J. Hum. Factors Ergon. Soc.* 57, 3 (May 2015), 407–434. DOI:https://doi.org/10.1177/0018720814547570

[19] Charles M. Judd, Laurie James-Hawkins, Vincent Yzerbyt, and Yoshihisa Kashima. 2005. Fundamental Dimensions of Social Judgment: Understanding the Relations Between Judgments of Competence and Warmth. *J. Pers. Soc. Psychol.* 89, 6 (2005), 899–913. DOI:https://doi.org/10.1037/0022-3514.89.6.899

[20] Nicolas Kervyn, Vincent Y. Yzerbyt, Charles M. Judd, and Ana Nunes. 2009. A Question of Compensation: The Social Life of the Fundamental Dimensions of Social Perception. *J. Pers. Soc. Psychol.* 96, 4 (2009), 828–842. DOI:https://doi.org/10.1037/a0013320

[21] Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey T. Hancock, and Michael S. Bernstein. 2020. Conceptual Metaphors Impact Perceptions of Human-AI Collaboration. In *Proceedings of the ACM on Human-Computer Interaction*, 26 pages. DOI:https://doi.org/10.1145/3415234

[22] Seo Young Kim, Bernd H. Schmitt, and Nadia M. Thalmann. 2019. Eliza in the uncanny valley: anthropomorphizing consumer robots increases their perceived warmth but decreases liking. *Mark. Lett.* 30, 1 (March 2019), 1–12. DOI:https://doi.org/10.1007/s11002-019-09485-9

[23] Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan. 2018. Algorithmic Fairness. *AEA Pap. Proc.* 108, (May 2018), 22–27. DOI:https://doi.org/10.1257/pandp.20181018

[24] Philipp Kulms and Stefan Kopp. 2018. A Social Cognition Perspective on Human–Computer Trust: The Effect of Perceived Warmth and Competence on Trust in Decision-Making With Computers. *Front. Digit. Humanit.* 5, June (2018), 1–11. DOI:https://doi.org/10.3389/fdigh.2018.00014

[25] Min Kyung Lee, Sara Kielser, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. 2010. Gracefully Mitigating Breakdowns in Robotic Services. In *Proceeding of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ACM Press, New York, New York, USA, 203–210. DOI:https://doi.org/10.1145/1734454.1734544

[26] Michal Luria, Rebecca Zheng, Bennett Huffman, Shuangni Huang, John Zimmerman, and Jodi Forlizzi. 2020. Social Boundaries for Personal Agents in the Interpersonal Space of the Home. *Conf. Hum. Factors Comput. Syst. - Proc.* (2020), 1–12. DOI:https://doi.org/10.1145/3313831.3376311

[27] Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *J. Soc. Issues* 56, 1 (January 2000), 81–103. DOI:https://doi.org/10.1111/0022-4537.00153

[28] Truong-Huy D Nguyen, Elin Carstensdottir, Nhi Ngo, Magy Seif El-Nasr, Matt Gray, Derek Isaacowitz, and David Desteno. 2015. Modeling Warmth and Competence in Virtual Characters. In *International Conference on Intelligent Virtual Agents*, Springer International Publishing, Cham, 167–180.

[29] Richard E. Nisbett and Timothy D. Wilson. 1977. The Halo Effect: Evidence for Unconscious Alteration of Judgments. *J. Pers. Soc. Psychol.* 35, 4 (1977), 250–256. DOI:https://doi.org/10.1037/0022-3514.35.4.250

[30] Raquel Oliveira, Patricia Arriaga, Filipa Correia, and Ana Paiva. 2019. The Stereotype Content Model Applied to Human-Robot Interactions in Groups. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE, 123–132. DOI:https://doi.org/10.1109/HRI.2019.8673171

[31] Rifca Peters, Joost Broekens, and Mark A. Neerincx. 2017. Robots Educate in Style: The Effect of Context and Non-verbal Behaviour on Children's Perceptions of Warmth and Competence. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, IEEE, 449–455. DOI:https://doi.org/10.1109/ROMAN.2017.8172341

[32] Pooja Prajod, Mohammed Al Owayyed, Tim Rietveld, Jaap Jan Van Der Steeg, and Joost Broekens. 2019. The Effect of Virtual Agent Warmth on Human-Agent Negotiation. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019)*, Montreal, Canada, 71–76.

[33] Seymour Rosenberg, Carnot Nelson, and P. S. Vivekananthan. 1968. A Multidimensional Approach to the Structure of Personality Impressions. *J. Pers. Soc. Psychol.* 9, 4 (August 1968), 283–294. DOI:https://doi.org/10.1037/h0026086

[34] Andrew Steain, Christopher John Stanton, and Catherine J. Stevens. 2019. The black sheep effect: The case of the deviant ingroup robot. *PLoS One* 14, 10 (October 2019), e0222975. DOI:https://doi.org/10.1371/journal.pone.0222975

[35] Adam Waytz and Michael I. Norton. 2014. Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking—not feeling—jobs. *Emotion* 14, 2 (2014), 434–444. DOI:https://doi.org/10.1037/a0036054

[36] W. Neil Widmeyer and John W. Loy. 1988. When You're Hot, You're Hot! Warm-Cold Effects in First Impressions of Persons and Teaching Effectiveness. *J. Educ. Psychol.* 80, 1 (1988), 118–121. DOI:https://doi.org/10.1037/0022-0663.80.1.118

[37] Bogdan Wojciszke, Roza Bazinska, and Marcin Jaworski. 1998. On the Dominance of Moral Categories in Impression Formation. *Personal. Soc. Psychol. Bull.* 24, 12 (December 1998), 1251–1263. DOI:https://doi.org/10.1177/01461672982412001

[38] Eva Yiwei Wu, Emily Pedersen, and Niloufar Salehi. 2019. Agent, Gatekeeper, Drug Dealer. *Proc. ACM Human-Computer Interact.* 3, CSCW (November 2019), 1–27. DOI:https://doi.org/10.1145/3359321

[39] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)*, ACM Press, Glasgow, Scotland, 1–12. DOI:https://doi.org/10.1145/3290605.3300509

[40] Vincent Yzerbyt, Valérie Provost, and Olivier Corneille. 2005. Not Competent but Warm... Really? Compensatory Stereotypes in the French-speaking World. *Gr. Process. Intergr. Relations* 8, 3 (July 2005), 291–308. DOI:https://doi.org/10.1177/1368430205053944

[41] Vincent Y. Yzerbyt, Nicolas Kervyn, and Charles M. Judd. 2008. Compensation Versus Halo: The Unique Relations Between the Fundamental Dimensions of Social Judgment. *Personal. Soc. Psychol. Bull.* 34, 8 (August 2008), 1110–1123. DOI:https://doi.org/10.1177/0146167208318602

A APPENDICES

A.1 Warmth and Competence Manipulations

		Warmth		
	Competence	House pricing	Car insurance	Movie recommendation
High	A **state-of-the-art artificial neural network algorithm** that was trained on data from **1,000,000 houses/car insurance plans/viewing choices and user ratings**.	A system that was developed to help **people like you** make better offers and achieve their goals.	A system that was developed to help **people like you** receive better car insurance offers and achieve their goals.	A system that is **friendly and well-intentioned**. The system's recommendation is based only on the information users provide, and it **does not promote specific content**.
Low	A **traditional decision tree algorithm** that was trained on data from **1,000 houses/car insurance plans/viewing choices and user ratings**.	A system that was developed to help **real estate agents** make better offers and achieve their goals.	A system that was developed to help **insurance agents** make better car insurance offers and achieve their goals.	A system that is **cold and exact.** The system's recommendation is based on the information users provide as well as some **promoted specific content**.

A.2 Pretests results

Manipulations were pre-tested to allow a calibration on high/low warmth/competence differences between descriptions. Participants were randomly assigned to read a description of one system (e.g., a low competence-high warmth system) and reported their perceptions of the system's warmth and competence (as described in studies 1A and 1C). Participants clearly distinguished between high/low warmth/competence levels. In the house pricing pretest, the high-warmth system was perceived as warmer (M = 4.74, SD = 1.26) than the low-warmth system (M = 4.03, SD = 1.48; $F(1, 130) = 8.74$, $p = .004$, Cohen's d = .51), and the high-competence system was perceived as more competent (M = 5.04, SD = 1.15) than the low-competence system (M = 4.61, SD = 1.29; $F(1, 130) = 4.11$, p = .045, Cohen's d = .35). Similar results were obtained in the movie recommendation pretest. The high-warmth system was perceived as warmer (M = 5.15, SD = 1.10) than the low-warmth system (M = 4.60, SD = 1.40; $F(1, 164) = 7.90$, $p = .006$, Cohen's d = .44), and the high-competence system was perceived as more competent (M = 4.98, SD = 1.20) than the low-competence system (M = 4.56, SD = 1.19; $F(1, 164) = 5.13$, $p = .025$, Cohen's d = .35). Note that in both pretests the competence condition did not affect the system's warmth perception and vice versa. There were also no interaction effects between the warmth and competence conditions on the system's warmth or competence perceptions.

A.3 Coronavirus study

In this study, participants were asked to choose one of two AI systems, which monitor coronavirus spread.

A.3.1 Method

Five hundred and ten participants (56.9% women, M_{age} = 32.56, SD = 10.88) were recruited via Prolific Academic, in return for monetary compensation. All participants were presented with the following prompt:

> In the last few months, the coronavirus has spread rapidly across the globe. Imagine that you decide to use an app that alerts people who have come in contact with an infected individual. The app is based on tracking the geographical location of its users and currently has over 100,000 users in your region.

Participants then learned that there are two apps in the market and read the description of the two apps, which were presented in random order. Participants were randomly assigned to one of five experimental conditions. In each condition, the apps varied in their levels of competence (high vs. low) and warmth (high vs. low). A high-competence app was described as an app that is *well established* and employs a *state-of-the-art artificial intelligence system*. The app's *location-identification accuracy* is within a radius of *0.5-meter* (1.5 feet) both indoors and outdoors, whereas a low-competence app as an app that is being *beta-tested* and employs *basic tracking tools*. The app's *location-identification accuracy* is within a radius of *100-meter* (330 feet) outdoors and might fail indoors. A high-warmth app was described as an app that is *friendly and well-intentioned*. It was *developed for people like them* to better protect themselves from potential infection, while a low-warmth app as an app that is *cold and exact*. It was *developed for governments* to better monitor citizens' movement and the spread of coronavirus. Similar to studies 1A-C, in four out of the five conditions, participants read one piece of information about each of the apps. That is, each one of the participants chose between two systems, one described in terms of warmth (high or low) and the other described in terms of competence (high or low). In the fifth condition, which was similar to studies 2B-C, participants chose between a high competence-low warmth app and a high warmth-low competence app. Finally, participants reported some demographics.

A.3.2 Results

Our manipulations were successful as more participants preferred the high-competence system (82.4%, N = 84) over the low-warmth system ($\chi^2(1) = 42.71$, $p < .001$) and the high-warmth system (65.7%, N = 67) over the low competence system ($\chi^2(1) = 10.04$, $p = .002$). No preference emerged between the low-warmth (53.9%, N = 55) and the low-competence systems ($\chi^2(1) = .63$, $p = .627$). Interestingly, more participants (63.4%, N = 64) chose the high-competence system over the high-warmth system ($\chi^2(1) = 7.22$, $p = .007$). similarly, more participants (64.7%, N = 64) chose the high competence-low warmth system over the high warmth-low competence system ($\chi^2(1) = 8.82$, $p = .003$).