Distribution Agreement

Signature:

________________________________ ________________

Shensheng Wang Date

Dimensions of Mind Perception in Faces

By

Shensheng Wang

DOCTOR OF PHILOSOPHY

Psychology

Philippe Rochat, Ph.D.	Patricia Bauer, Ph.D.
Advisor	Committee Member
Scott O. Lilienfeld, Ph.D.	Daniel D. Dilks, Ph.D.
Committee Member	Committee Member
Robert McCauley, Ph.D.	Yunxiao Chen, Ph.D.
Committee Member	Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.

Dean of the James T. Laney School of Graduate Studies

Date

Dimensions of Mind Perception in Faces

By

Shensheng Wang

M.A., Emory University

Advisor: Philippe Rochat, Ph.D.

An abstract of

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Psychology

2019

Abstract

Dimensions of Mind Perception in Faces

By

Shensheng Wang

Philosophers have long argued that we do not have a direct, perceptual, access to the mental lives of others. Nevertheless, recent development in philosophy and psychology challenges this epistemological position by raising the questions of whether we *perceive* minds in faces and if so, what perceptual experience it entails. In four studies, I addressed these questions by examining whether faces automatically trigger mind awareness, and particularly what minds human faces elicit compared with artificial (Study 1 and Study 2) and dog faces (Study 3). In addition, I probed the cognitive penetrability of mind perception by examining how knowledge about others' mental capacities influences the perception of minds in faces (Study 4). The studies showed that compared with human faces, both artificial and dog faces were automatically imbued by participants with lesser minds, in other words, dehumanized. In particular, artificial and dog faces elicited differential dehumanization, the magnitude of which varied along two distinct dimensions—agency and experience. Furthermore, prior beliefs about someone lacking either agency or experience lowered participants' threshold for detecting minds in faces. Taken together, these findings corroborate the two-dimensional structure of mind perception in faces and provide preliminary evidence for our perceptual knowledge of other minds via face perception. I discuss the implications of these findings for the *other minds problem* in philosophy and the *uncanny valley* phenomenon in robotics. In closing, I point to tracing the developmental roots of mind perception in infant face perception as a fruitful future direction.

Dimensions of Mind Perception in Faces

By

Shensheng Wang

M.A., Emory University

Advisor: Philippe Rochat, Ph.D.

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Psychology

2019

Acknowledgements

I am grateful for the tireless support of my graduate advisor, Philippe Rochat. His guidance was invaluable.

Further, I thank other members on my committee, Scott Lilienfeld, Robert McCauley, Patricia Bauer, Daniel Dilks, and Yunxiao Chen, for their thoughtful comments on this project and beyond.

I also want to thank other professors who taught me and gave me helpful advice: Yuk Fai Cheong, Phillip Wolff, Stella Lourenco, Gregory Berns, Lynne Nygaard, and Dietrich Stout.

I would also like to thank my colleagues: Erin, Ginger, Kit, Theresa, Sara, Cynthia, Maria, Natalie, Sean, and Joon, and friends: Dennis, Divya and Daniel, Robert, Darrell, Carlos and Fernanda.

Table of Contents

Dimensions of Mind Perception in Faces

How humans perceive minds in others is a fundamental question of particular interest to philosophers and psychologists studying social cognition (Dennett, 1996). Philosophical discussion tends to emphasize whether humans have a direct access to others' minds (Varga, 2017), whereas psychology research tends to focus on the conditions under which humans attribute minds to (i.e., mind perception; Wegner, 2002) or withhold minds from other people (i.e., dehumanization; Haslam, 2006).

The last decade and half has witnessed an increasing integration between these two intellectual traditions (Ishiguro, 2006; Kwan & Fiske, 2008; Varga, 2017; Waytz, Epley, & Cacioppo, 2010), whereby converging evidence suggests that we might directly *perceive* minds in faces (Fincher & Tetlock, 2016; Looser & Wheatley, 2010). Nevertheless, scientific progress in understanding how we perceive minds in faces is hindered by disagreements regarding what it means to perceive "minds," which leaves the question open regarding whether our knowledge of other minds has a genuine perceptual nature rooted in face perception.

To address this question, I adopt a novel approach to examining the face-mind linkage by elucidating the dimensions on which humans perceive minds (and the lack thereof) in faces. To begin with, I identify two primary approaches in the existing literature—Mind Addition vs. Mind Subtraction—which I argue rest upon distinct presumptions about what constitutes the perception of minds. Next, I discuss the differences between these two approaches and point to the challenges they raise for understanding the face-mind linkage. I then briefly overview research on anthropomorphism and dehumanization, based on which I report four current studies, in which I validated the well-documented two-dimensional model of mind perception in faces. In closing, I discuss the implications and limitations of the present findings and suggest future directions.

A Brief History of Research on Mind Perception

The peculiar ability of humans to perceive fellow people as minded and to explain and predict their behavior in terms of desires, intentions, and beliefs is arguably the most distinctive feature of humanity. Although none of us is alone in the universe to possess a mind (i.e., the philosophical idea of solipsism), we seem limited in our ability to reason about minds other than our own. This particular challenge we face in our daily life has to do with the uncertainty of knowing whether other minds exist, an epistemological conundrum known as the *other minds problem*. In other words, the problem of other minds is the question about how we come to be aware that other people possess minds like us.

Philosophers and cognitive scientists try to solve this problem by focusing primarily on the *cognitive processes* by which humans infer the mental states of others, a cognitive ability commonly known as theory of mind (ToM, Premack & Woodruff, 1978), mentalizing (Frith & Frith, 2003) or mindreading (Gallese & Goldman, 1998; Heyes & Frith, 2014). Two of the most prominent theories in the literature on ToM include theory theory (TT) and simulation theory (ST). TT posits that we understand others' minds by forming quasi-scientific theories to explain correlations between mental states and behavior (Gopnik & Meltzoff, 1997; Wellman, 1992). In contrast, ST posits that humans rely on their own minds as an "off-line" model to gain insights of others' minds via mental simulation—the reenactment of others' mental lives (Goldman, 2013;

Gordon, 2008). Despite the apparent opposition between TT and ST, they seem to agree on one fundamental assumption—it is *not* possible to have a direct, perceptual, grasp of others' mental lives without the involvement of cognitive processes. Contrary to this assumption, phenomenologists argue that perception alone might provide us with insights about other minds (Gallagher, 2008), although the extent to which we directly perceive the desires, thoughts, and intentions remains an open question (Gallagher & Varga, 2014; Grossmann, 2017).

Philosophers help clarify this issue by drawing a distinction between *mind awareness* (i.e., to perceive others as minded creatures with potentials to think, act, and interact with our own) and *mental state awareness* (i.e., ToM or to recognize what states the minds might be in). Literary translator Duddington argues, "we recognize the presence of other minds in the same direct and immediate fashion as we do the presence of bodies" (Duddington, 1918, p. 166). Philosopher Wittgenstein echoes, "Whether a body is minded is something that one perceives through the senses" (Varga, 2017). In light of this distinction, philosophers argue that it is possible to have a direct, perceptual, access to mind awareness independent of mental state awareness (Varga, 2017).

However, psychologists traditionally studying social cognition, particularly, *mind perception,* often neglect to consider this crucial distinction, paying little attention to what might be the perceptual roots of the attribution of minds to humans and nonhuman agents (Epley & Waytz, 2010; Wegner, 2002) Perhaps the only exception to this claim concerns a growing body of literature on *face animacy perception* (Looser & Wheatley, 2010). This emerging field of research inherits much from two traditionally separate yet increasingly integrated lines of research, namely, mind perception in social psychology

(Wegner, 2002) and face perception cognitive psychology and neuroscience. In what follows, I review the relevant literature by identifying the fundamental differences between these two approaches and point to the theoretical and methodological issues they raise.

Two Approaches to Examining the Face-Mind Linkage

Mind Addition. Originating from a social psychology tradition in mind perception (Wegner, 2002), the Mind Addition approach focuses on factors that contribute to the attribution of minds to faces. Within this tradition, research demonstrates that both the human-likeness (Looser & Wheatley, 2010) and configural processing of faces (i.e., the processing of facial features as their relations among each other instead of the features in isolation by themselves; Deska, Almaraz, & Hugenberg, 2017; Deska & Hugenberg, 2017; Hugenberg et al., 2015) contribute to the ascription of minds to faces. In addition, group membership is shown to modulate the attribution of minds to faces, whereby outgroup faces are imbued with lesser minds compared with ingroup faces, and this effect persists even if group memberships are arbitarily defined (Hackel, Looser, & Van Bavel, 2014; Hackel, Mende-Siedlecki, Looser, & Van Bavel, 2015; Powers, Worsham, Freeman, Wheatley, & Heatherton, 2014; Swiderska, Krumhuber, & Kappas, 2013).

A common means by which researchers measure the attribution of minds to faces is by asking participants to rate the perceived animacy (e.g., how alive does the face look?) or mind (e.g., does the face possess a mind?) of faces on a Likert-type scale (e.g., 1 = definitely alive, 7 definitely not alive). For example, in their seminal work, Looser and Wheatley (2010) used image morphs created from a doll face and a human face to examine the minimal visual cues for a face to be judged alive or possessing a mind. They found that the point of subjective equality (PSE), the point at which a face is equally likely to be deemed alive and not alive, consistently lies at a location near the human endpoint of the morph continuum (see Figure 2, Looser & Wheatley, 2010, p. 1856). *Mind Subtraction.* In contrast, stemming from a cognitive psychology and neuroscience tradition in face perception, the Mind Subtraction approach focuses on the perceptual discrimination between human faces and artificial faces (e.g., doll faces; Balas & Koldewyn, 2013; Farid & Bravo, 2012; Looser, Guntupalli, & Wheatley, 2013; Wheatley, Weinberg, Looser, Moran, & Hajcak, 2011). Within this tradition, researchers demonstrate behavioral and neural responses distinguishing animate (e.g., human faces) from inanimate faces (e.g., doll faces) (Looser et al., 2013; Wheatley et al., 2011). For example, researchers found that although both animate and inanimate faces elicit face-specific electrophysiological responses (N170/VPP), only animate faces elicit sustained late positive potential (LPP) beyond 400ms following stimulus onset, demonstrating a later stage in face processing discriminating between animate from inanimate faces (see Figure 2, Wheatley et al., 2011, p. 3). The authors referred to this discrimination as the perception of minds in faces.

These two approaches raise fundamental questions regarding the conception and measurement of minds in faces: First, what does it mean to perceive minds in faces? Does it entail the attribution or the discrimination of minds to faces? Second, how do we measure the different minds people perceive in faces? Do they differ in kind (i.e., presence vs. absence of minds) or in degrees (i.e., more or less minds)? Finally, how can we study the perception of minds in faces? Should minds be measured on a continuous scale as a dependent variable or operationally defined in terms of the contrast between human (animate) and artificial (inanimate) faces as an independent variable? Which of these different conceptions and measurements can provide more accurate insights into the perception of minds in faces?

To address these questions, I consider the null hypothesis, against which each approach defines what constitutes the perception of minds. It is the null hypothesis underlying each approach that made me to label them as Mind Addition and Mind Subtraction, respectively.

The Mind Addition approach (Deska et al., 2017; Looser & Wheatley, 2010) defines the perception of minds in faces against the null hypothesis that observers initially assume no minds to be present in the face, unless the face proves otherwise[1] by eliciting, for example, perception of human characteristics in it. Therefore, according to the Mind Addition approach, a process of "adding" minds to "mindless" faces constitutes the perception of minds in faces. For example, although you know that a robot is not alive, but when the humanlike face of a robot prompts you to wonder whether it has the potential to feel pain, you may attribute a mind to it (K. Gray & Wegner, 2012). Similarly, simple geometric shapes chasing each other can elicit the attribution of intentionality to them, in part, because perceivers presume that geometric shapes, when static or moving in a random pattern, are objects, which should not possess minds (Gao & Scholl, 2011). However, when it begins following another object, it leads you into thinking that it must possess humanlike intentionality to account for this behavior.

[1] One way to conceive this presumption is to consider an analogy—a face is mindedless until proven minded in the same way as a person is innocent until proven guilty.

In contrast, the Mind Subtraction approach defines the perception of minds in faces against the null hypothesis that a face possesses a mind by default, until it proves otherwise by revealing, for example, physical features that violate our domain-specific expectations for the prototypical human face. Therefore, according to the Mind Subtraction approach, a process of “subtracting” minds from “minded” faces constitutes the perception of a lack of minds in faces. For example, at first sight, you might perceive Giuseppe Arcimboldo’s painting *Fruit Basket* as a face belonging to an affluent businessman, who definitely possesses a mind; however, as you soon recognize that the facial features (e.g., eyes, nose, and cheeks) are composed of different types of fruits, you come to realize that it should not possess a mind.

Despite the previously mentioned differences between the two approaches, they do agree on one assumption, that is, they both construe “minds” as a unitary construct—Faces can have minds that differ either in kinds (e.g., presence vs. absence of minds) or in degrees (e.g., possessing more or less minds), but the perceived mind in faces always varies along a single dimension. This assumption, however, runs counter to findings in the extent literature on mind perception and dehumanization, which clearly suggests that humans perceive minds along two distinct dimensions (Fiske, Cuddy, & Glick, 2007; H. M. Gray, Gray, & Wegner, 2007; Haslam, 2006).

The inclination to not consider minds as a multidimensional construct may be attributable in part to the failure of earlier research to demonstrate their dissociation in the attribution of minds to faces. In particular, Looser and Wheatley (2010) found that regardless of whether participants are asked to judge faces on perceived animacy, or the ability to either formulate plans (agency), feel pain (experience), or possess a mind, the point of subjective equality (PSE) remains the same (see Figure 2, Looser & Wheatley, 2010, p. 1856). In addition, the ratings of these attributes were highly intercorrelated. Therefore, although these findings support the interchangeability between mind and animacy (Pearson's r = .922; Santos, David, Bente, & Vogeley, 2008), they provide little evidence for the multidimensionality of mind perception in faces (Pearson's *r*s = .958 and .953 for ratings of agency and animacy and ratings of experience and animacy). Given that multidimensionality is arguably a defining feature of what we mean by mind, the current literature on face animacy perception leaves the question open regarding whether we indeed perceive faces as attached to "minds" (Looser & Wheatley, 2010).

Philosophers long warn against the danger of confusing ontological questions (*what exists*) with epistemological ones (*where knowledge comes from*). In this case, however, addressing the question of *how* we gain mind awareness necessitates elucidating *what* perceiving minds in faces entails.

Therefore, my dissertation sets out to address the *epistemological* question of whether we perceive minds in faces by first solving the *ontological* question of what dimensions of minds in faces humans perceive. Given the failure of face perception research to corroborate the multidimensionality of mind, I borrowed theories and methodologies from social psychology research on anthropomorphism and dehumanization, to reexamine the dimensional structure of mind perception in faces. Next, I briefly review this literature, based on which I conducted four studies to validate the well-documented two-dimensional model of mind perception in faces.

Anthropomorphism and Dehumanization

Anthropomorphism is a prevalent phenomenon that pertains to the attribution of human characteristics to a variety of nonhuman entities, including nonhuman animals (Eddy, Gallup, & Povinelli, 1993), mechanical devices (Duffy, 2003; Vidal, 2007), supernatural agents (Barrett, 2000; Barrett & Keil, 1996; Guthrie, 1993), and even the world (Banerjee & Bloom, 2014). According to David Hume (Hume, 1757/1957, p. 29), “there is a universal tendency among mankind to conceive all beings like themselves.” In the psychological literature, researchers have used the term “anthropomorphism” to broadly describe processes that entail attributing human characteristics, in particular, a human mind, to nonhuman and nonliving entities.

Dehumanization, in contrast, entails perceiving humans as nonhuman entities (e.g., animals or machines) and denying them human characteristics (Haslam & Loughnan, 2014). Pioneering research on dehumanization was conducted in social psychology in the 1970s with a focus on violence (Kelman, 1976; Staub, 1989). Since the early 1990s, researchers have shifted the conceptualization of dehumanization, extending research from extreme (e.g., genocide) to subtler forms of dehumanization (e.g., stereotyping and prejudice).

In particular, researchers demonstrate that there is a subtle form of dehumanization called *infrahumanization*, whereby the denial of uniquely human characteristics to outgroup members occurs without involving physical violence or aggression (Leyens et al., 2001). One useful paradigm for studying this phenomenon is the implicit association test (IAT), which involves responding to both target (e.g., ingroup and outgroup members) and attribute categories (e.g., human and animal words) simultaneously across multiple blocks. In some blocks (congruent blocks), participants are required to use the same key to respond to both ingroup members and human words, whereas in others (incongruent blocks), participants are required to respond to both ingroup members and animal words (see Appendix A for a trial from a congruent block in Study 1). By showing that there is a stronger association, indexed by higher accuracy and shorter reaction time in participants' responses, between ingroup (outgroup) members and human (animal) words relative to the opposite target-attribute association, researchers reliably demonstrate the denial of human characteristics to racial outgroups (Leyens, Demoulin, Vaes, Gaunt, & Paladino, 2007; Leyens et al., 2001; Paladino et al., 2002), women (Viki & Abrams, 2003), and animals (Bilewicz, Imhoff, & Drogosz, 2011) in a subtle form.

Despite the fact that these two phenomena have primarily been studied separately, converging evidence has shown that humans perceive mind and humanness similarly in both humans and nonhuman entities. According to the *Mind Perception* theory, mind perception broadly entails attributing humans' mental capacities on two distinct dimensions: the ability to plan and act (*agency*) and the ability to feel and experience (experience; H. M. Gray et al., 2007). Because the dimensions of agency and experience capture distinct human mental capacities, agents can be perceived as both high on agency and relatively low on experience (e.g., God) or both low on agency and relatively high on experience (e.g., infants).

Likewise, the Dual Model theory of dehumanization (Haslam, 2006) posits that humanness entails two distinct meanings—Human Uniqueness (HU) and Human Nature (HN). Uniquely human characteristics such as language and refined emotions (e.g., shame and pride) reflect cultural learning and socialization and are therefore expected to vary across cultures. Human Uniqueness defines the categorical boundary between humans and nonhuman animals, whereby denying human uniqueness (i.e., *animalistic dehumanization*) renders people to be perceived as lacking self-control, intelligence, and rationality and less cultured. In contrast, Human Nature entails characteristics such as emotionality and warmth, which are inherent and central to all humans, whereby denying human nature (i.e., *mechanistic dehumanization*) renders humans to be perceived as lacking warmth and emotions.

Converging evidence suggests that the ways in which we dehumanize persons are inextricably connected with the ways in which we anthropomorphize nonhuman agents, such that there may exist a continuum between anthropomorphized robots and dehumanized people (Haslam, 2006; Kwan & Fiske, 2008; Waytz et al., 2010). For example, research shows that both anthropomorphism and dehumanization are modulated by top-down (e.g., motivation) and bottom-up (e.g., visual cues of a face) variables. Given the link between anthropomorphism and dehumanization and the fact that both experience and Human Nature reflect the distinctions between humans and machines, whereas both agency and Human Uniqueness reflect the distinctions between humans and other animals, I refer to them interchangeably as the two dimensions of mind perception–*experience (Human Nature)* and *agency (Human Uniqueness)*.

Current Research

In the current research, I focus on elucidating the dimensions along which people perceive minds in faces. To do this, I examined the extent to which the perceptual discrimination between animate (human) and inanimate (artificial) faces map onto the two forms of dehumanization—animalistic and mechanistic. In particular, if the perceptual discrimination between artificial and human faces entails perceiving different minds in faces, artificial faces should be associated with animal words more closely than with human words, that is, dehumanized. To test this, in Study 1, I adapted the IAT paradigm used in research on infrahumanization. I predicted that participants would "dehumanize" artificial faces by demonstrating a stronger association between human (artificial) faces and human (animal) words relative to the opposite target-attribute association, which would suggest that artificial faces elicit (animalistic) dehumanization.

Once establishing that artificial faces are dehumanized (on the human uniqueness dimension), in Study 2, I next examined whether artificial faces are dehumanized also on the human nature dimension (i.e., mechanistic dehumanization) by using human and automata words to reflect the two opposing attribute categories in the IAT paradigm. Based on previous findings that robots are perceived as intrinsically lacking experience (e.g., the capacity to feel; K. Gray & Wegner, 2012), I predicted that artificial faces would be dehumanized more strongly on the Human Nature dimension, a finding which would support the dissociation between the two dimensions of mind perception in faces.

In the dehumanization literature, the distinction between humans and nonhuman animals is the basis of Human Uniqueness. Therefore, the contrast between human faces and faces of nonhuman animals are as critical for understanding the perceptual basis of mind perception in faces as the contrast between human and artificial faces. Nevertheless, for historical reasons, the majority of research on face animacy perception relies on artificial faces (e.g., dolls and computer-generated characters) or image morphs created from them to examine the perception of minds in faces. To remedy this omission, in Study 3, I included nonhuman animal (e.g., dog) faces to examine the extent to which they are dehumanized and particularly focused on how the two forms of dehumanization might manifest differently between artificial faces and dog faces. I predicted that unlike artificial faces, which are more strongly dehumanized on the Human Nature dimension, animal faces should elicit dehumanization on both dimensions, but more so on the Human Uniqueness dimension.

Finally, in Study 4, I probed the cognitive penetrability of mind perception in faces. Provided that dehumanization is the opposite process of anthropomorphism, I argue that factors increasing participants' propensity to dehumanize others should decrease the propensity to perceive minds in faces. To test this, I presented the same participants with three different kinds of vignettes describing protagonists as either losing their mental capacities to feel (Loss of Experience), to plan (Loss of Agency), or fully recovered from a car accident (Control) by using a within-subject design. In each condition, I measured the threshold for perceiving minds in faces along a morph continuum. The prediction was that compared with the control condition, knowing the protagonists' loss of mental capacities would decrease participants' propensity to perceive minds in faces, thereby shifting the threshold toward the doll endpoint, which would suggest a top-down effect. Given that artificial faces are dehumanized more on the Human Nature dimension, I predicted that lacking experience would yield a more pronounced top-down effect.

Study 1

In Study 1, I adapted the IAT paradigm to examine whether artificial faces would elicit dehumanization, given that people readily detect the lack of minds of artificial faces compared with human faces. To do so, I replaced the target categories (i.e., ingroup and outgroup members) in Viki et al. (2006) with human and artificial faces and used human and animal words to represent the opposing attribute categories. In particular, I included three separate sets of stimuli of human and artificial faces to examine the robustness and generalizability of the findings.

Because the IAT methodologies achieve higher efficacy when items are representative of their target or attribute categories, I selected faces that are clearly inanimate (e.g., the face of a computer-generated character) or animate (e.g., the face of its human inspiration). Participants categorized faces as either artificial or human, yet they also simultaneously categorized stimulus words as related to either human or animal. I predicted that across the stimulus sets, participants would more strongly associate human faces with human than with animal words and associate artificial faces more strongly with animal than with human words.

Methods

Participants

Forty-two undergraduate students of Emory University (*Mage* = 19.4, *SD* = 1.2, 24 Males) participated in this study. Data from 3 additional participants were deleted listwise due to missing data. The sample size was determined in accordance with Study 1 in Viki et al. (2006). Participants were debriefed and received course credits after completing the study.

Stimuli

Words. Ten human words and 10 animal words were adapted from Viki et al. (2006)[2]. Words are matched on their overall valence and differ significantly in their level of human uniqueness, human words rated as significantly more uniquely human compared with animal words (see Appendix B for supplementary materials).

Faces. Thirty artificial faces and 30 human faces were divided into three stimulus sets: Sims, LW, and UV, each tested in a separate IAT.

Sims: Computer-generated images (CGIs) of 10 young, female, Caucasian celebrities adapted from the computer game "Sims 4" and the photos of their human inspirations were collected by conducting a Google search. The selected images were well-matched in terms of their identities.

LW: Face images of 10 mannequins and 10 humans were adapted from Looser and Wheatley (2010). The selected images were well-matched in terms of their gender, race, and age, but not identities.

UV: Ten artificial faces eliciting eerie feelings and 10 human faces were adapted from our previous study examining participants' affective responses to human replicas with varying degrees of human-likeness (Wang & Rochat, 2017).

[2] Human words include Wife, Maiden, Woman, Person, Husband, Humanity, People, Civilian, Man, and Citizen. Animal words include Pet, Mongrel, Pedigree, Breed, Wildlife, Critter, Cub, Creature, Feral, and Wild.

Design and Procedure

The experiment was introduced as a reaction time study in social psychology. Participants completed three IATs in randomized order on computers under the supervision of an experimenter. Following the standard procedure (Greenwald, McGhee, & Schwartz, 1998), each IAT consisted of 7 blocks.

During each IAT, participants first practiced sorting faces (Block 1, 20 trials) and words (Block 2, 20 trials) into their respective categories in a single categorization task using two response keys (E and I on the keyboard). Participants then practiced sorting both faces and words simultaneously in a combined categorization task (Block 3, 20 trials) before they completed the actual task of the same nature (Block 4, 40 trials). During the combined categorization task, faces and words were assigned to the same keys participants had previously used to sort stimuli in Block 1 and 2. In Block 5, however, participants learned to sort words with response keys reversed. Following the recommendation of Nosek, Greenwald, and Banaji (2005), I doubled the length of this block, resulting in 40 practice trials. The last two blocks (i.e., Block 6 and 7) replicated Block 3 and 4, except that participants sorted faces and words with the newly acquired response pattern. Therefore, each IAT consisted of both a compatible block, where participants used the same keys to respond to human (artificial) faces and human (animal) words, and an incompatible block, where participants used the same keys to respond to human (artificial) faces were paired with animal (human) words. The design was counterbalanced such that half of the participants completed the compatible block first and the other half completed the incompatible block first.

On each trial, the stimulus was presented vertically and horizontally centered (see Appendix A) on the screen and remained on screen until a response was made or until 3000 ms has elapsed. Participants were instructed to provide each response as quickly and as accurately as possible. If their initial response was incorrect, a red letter *X* appeared, and participants were instructed to give the correct response after seeing the red letter *X*. The red letter *X* remained on screen until the participant made the correct response. An inter-stimulus interval of 400 ms followed each correct response. Stimuli were presented in a randomized order and without replacement until the available stimuli for each block were exhausted. Each stimulus appeared once in each block. The experiment was presented in the web browser by using iatgen (Carpenter et al., 2018).

Planned Data Analysis

Data analysis followed the protocol outlined by Greenwald, Nosek, and Banaji (2003). All trials of the two combined categorization tasks were analyzed. Trials with reaction times longer than 10000 ms and participants who had more than 10% of trials faster than 300 ms were excluded. The remaining trials of the remaining participants entered subsequent data analysis. A *D* measure was calculated by dividing the differences between mean reaction times for the compatible and incompatible tasks ($RT_{compatible} – RT_{incompatible}$) by the standard deviation of RT in trials from both tasks (Nosek et al., 2005). After computing a *D* measure for each participant, a one-sample t–test determined whether the sample *D* measures was statistically significant from 0, which indicates a difference in the strength of implicit association between compatible and incompatible pairings. Here, a positive *D* measure suggests a stronger association between artificial faces and animal words, and between human faces and human words, compared with the reverse pairings.

Results and Discussion

Because a D score of zero indicates an absence of association between faces and words whereas a positive D score indicates a predicted effect of face dehumanization, D scores from the three IATs were first compared with zero to examine whether artificial faces were dehumanized and then compared between each other to examine the robustness of the findings across the three stimulus sets.

Are artificial faces dehumanized? To answer this, I first conducted a one-sample t test comparing mean D scores with zero separately for each IAT. Results showed that consistent across all three stimulus sets, D scores were significantly greater than 0: Sims (M = .74), t(41) = 18.77, p < .0001, 95% CI[.67, .82], LW (M = .67), t(41) = 14.60, p < .0001, 95% CI[.58, .75], and UV (M = .67), t(41) = 15.66, p < .0001, 95% CI[.58, .74].

Are artificial faces differentially dehumanized across stimulus sets? A one-way repeated-measures analysis of variance (ANOVA) comparing the mean D scores between the three IATs showed that D scores from each IAT were not significantly different from each other, F(2, 82) = 1.12, p = .33, η_p^2 = .03. None of the pairwise comparisons reached statistical significance. Results are summarized in Figure 1.

In conclusion, Study 1 demonstrates that humans tend to automatically associate artificial faces more strongly with animal words than with human words. This finding links perceived face animacy to a subtle form of dehumanization, i.e. the likening of artificial faces to nonhuman animals possessing less uniquely human attributes, such as higher levels of cognition, moral sensitivity, and sophistication. In this respect, perceived lack of face animacy shares the same mechanisms with infrahumanization in intergroup relation in that both forms of dehumanization entail the denial of Human Uniqueness (HU)[3]. Nevertheless, it remains unknown whether artificial faces might be similarly dehumanized on the Human Nature (HN) dimension of humanness.

In fact, research on the uncanny valley hypothesis points to experience, which is akin to Human Nature, as a fundamental distinction between humans and androids such that the perception of experience in androids violates people's expectations for automata, eliciting eerie feelings in humans, thereby creating the uncanny valley phenomenon (K. Gray & Wegner, 2012). To address this question, I conducted Study 2 to examine the implicit association between human and artificial faces with a novel attribute category—automata.

Study 2

In Study 2, I examined whether artificial faces might be additionally dehumanized on the Human Nature dimension by using human and automata words to reflect the two

[3] One might argue that artificial faces cannot be targets of dehumanization, because by definition they are not human. Nevertheless, I would argue that first this claim assumes a rather narrow definition of dehumanization, which neglects the link between mind perception and dehumanization. Given their link, researchers construe dehumanization "not as an exceptional process that must be understood only on its own terms but rather as a phenomenon that is related to the fundamental processes of mind attribution." (Haslam & Loughnan, 2014, p. 404). Therefore, by arguing that the differential association between human (artificial) faces and human (animal) words demonstrates the dehumanization of artificial faces, I adopt this broader definition of dehumanization in terms of the attribution of lesser minds to other humans and nonhuman entities.

opposing attribute categories. Furthermore, I compared the magnitude between the two forms of dehumanization—animalistic and mechanistic— across the three stimulus sets. Given the previous findings that people tend to perceive androids as fundamentally lacking the capacities for experience compared with humans, I predicted that for artificial faces, mechanistic dehumanization (i.e., denying Human Nature) would be stronger than animalistic dehumanization (i.e., denying Human Uniqueness).

Methods

Participants

Seventy-nine undergraduate students of Emory University (*Mage* = 19.3, *SD* = 2.3, 34 Males) participated in this study[4]. Data from 14 additional participants were deleted listwise due to missing data. All participants were debriefed and received course credits after completing the study.

[4] The sample size was originally planned to examine the potential cultural differences in mind perception in faces, given the different emphasis Anglo-Australian and ethnic-Chinese participants place when perceiving outgroups along the two dimensions of humanness (Bain, Park, Kwok, & Haslam, 2009). Nevertheless, due to the small sample of participants from East Asia (N = 17), I did not conduct this analysis.

Stimuli

Words. Ten animal and 10 automaton words were adapted from Loughnan and Haslam (2007)[5] (see Appendix C for Supplementary materials), and 10 human words were the same as those used in Study 1.

Faces. Faces were the same as those used in Study 1.

Design and Procedure

The experiment was introduced as a reaction time study in social psychology. Participants completed six IATs in a randomized order on computers in small group under the supervision of an experimenter. Each IAT followed the same 7-block procedure as outlined in Study 1. The experiment was presented in the web browser by using iatgen (Carpenter et al., 2018).

Planned Data Analysis

Data analysis followed the same procedure in Study 1.

Results and Discussion

Are artificial faces dehumanized on both HU and HN dimensions? For all of the six IATs, D scores were significantly greater than zero, suggesting that artificial faces were dehumanized on both HU and HN dimensions and across the three stimulus sets, *ps* < .001. Results are summarized in Table 1.

[5] Animal words include Alligator, Animals, Beast, Cattle, Chimpanzee, Elephant, Kangaroo, Mammals, Platypus, and Primates. Automata words include Android, Artificial, Automaton, Computer, Laptop, Machine, Mechanical, Robot, Software, and Synthetic.

On which dimension are artificial faces more strongly dehumanized? A 2 (Attribute: animal vs. automata) x 2 (Stimulus Set: Sims, LW, and UV) repeated-measures analysis of variance (ANOVA) yielded significant main effects of stimulus set, $F(2, 156) = 3.39$, $p = .04$, $\eta_p^2 = .04$, and attribute type, $F(1, 78) = 11.30$, $p = .001$, $\eta_p^2 = .13$, but did not yield any significant interaction between the two, $F(2, 156) = .38$, $p = .68$, $\eta_p^2 = .005$. Paired t tests further corroborated the finding that mechanistic dehumanization was significantly stronger than animalistic dehumanization for all stimulus sets, but LW. Results are summarized in Table 2 and Figure 2.

These findings suggest that artificial faces are perceived as lacking not only human nature but also human uniqueness, therefore eliciting both forms of dehumanization (Haslam, 2006). Furthermore, for two of three stimulus sets, artificial faces were dehumanized more strongly on the Human Nature than on the Human Uniqueness dimension, corroborating the two-dimensional model of mind perception in faces.

In Study 1 and 2, I demonstrate that compared with human faces, artificial faces elicit both animalistic and mechanistic dehumanization. These findings suggest that face animacy entails perceiving both dimensions of humanness (i.e., human uniqueness and human nature) in faces. In this respect, the current findings reject the existing conceptualization of perceived face animacy as a unitary construct (Looser & Wheatley, 2010) and as a rigid dichotomy (Wheatley et al., 2011). The current findings contribute to the literature on mind perception by demonstrating that perceiving minds in nonhuman faces is similar to attributing minds to nonhuman agents (H. M. Gray et al., 2007), which both entail perceiving other minds along two distinct dimensions. The current findings also provide a novel paradigm for examining face animacy perception based on implicit social cognition. This novel methodological approach is complementary to the existing methods, which include the use of image morphs and the measurement of brain responses toward human versus doll faces. This novel paradigm allows researchers to reveal the multidimensional nature of mind perception in faces, which is otherwise hidden by previous research.

Study 3

ERP research provides conflicting evidence regarding the neural organization of the animacy and species of faces. Whereas Looser et al. (2013) demonstrate that the discrimination between dog and human faces (i.e., face species) is based on differences in global face form and not based on the degree of mind, accumulating evidence points to the shared global structure of face across species (Balas & Koldewyn, 2013; Rousselet, Macé, & Fabre-Thorpe, 2004). In particular, Balas and Koldewyn (2013) found that the P100, an early face-specific ERP component, is sensitive to both face species and face animacy. The conflicting findings raise a question regarding whether the contrast between human and nonhuman animal faces reflects the differences in global form (e.g., "faceness"; Shannon, Patrick, Venables, & He, 2013) or in mind.

In Study 3, I addressed this question by examining the extent to which animal faces, like artificial faces, similarly elicit two forms of dehumanization, and if so, on which dimension they are more strongly dehumanized. I predicted that animal faces would similarly elicit dehumanization on both dimensions, but would be more pronounced on the HU dimension, given that animal faces activate the implicit association between faces and nonuniquely human traits, which distinguish humans from nonhuman animals.

Methods

Participants

Forty-three undergraduate students of Emory University (*Mage* = 20.0, *SD* = 3.99, 21 Males) participated in this study. Data from 5 additional participants were deleted listwise due to missing data. I collected as many participants as possible before the end of the semester. Participants were debriefed and received course credits after completing the study.

Stimuli

Words. Ten animal and 10 automaton words were the same as those used in Study 2.

Faces. Ten artificial and 10 human faces were from the UV stimulus set, which were used in both Study 1 and Study 2. Ten dog faces (huskies) were downloaded and selected from Flickr.

Design and Procedure

The experiment was introduced as a reaction time study in social psychology. Participants completed four IATs in a randomized order on computers in small group under the supervision of an experimenter. Each IAT followed the same 7-block procedure as outlined in Study 1. The experiment was presented in the web browser by using iatgen (Carpenter et al., 2018).

Planned Data Analysis

Data analysis followed the same procedure in Study 1.

Results and Discussion

Are dog faces dehumanized on both HU and HN dimensions? For IATs that used either artificial or dog faces, D scores were significantly greater than zero, suggesting that similar to artificial faces, dog faces were dehumanized on both HU and HN dimensions, *ps* < .001. Results are summarized in Table 3.

On which dimension are dog and artificial faces more strongly dehumanized? A 2 (Target: dog vs. artificial faces) x 2 (Attribute: animal vs. automata words) repeated-measures analysis of variance (ANOVA) yielded a significant interaction effects of target type and attribute type, $F(1, 42) = 7.54$, $p = .009$, $\eta_p^2 = .15$. Paired t tests showed that dog faces elicited animalistic dehumanization more strongly than mechanistic dehumanization, whereas artificial faces showed the opposite pattern, although the differences for artificial faces was not statistically significant. Results are summarized in Table 4 and Figure 3.

In Study 3, I found that not only artificial faces but also animal faces (dog faces in this study) are dehumanized on both HU and HN dimensions. Importantly, dog faces are dehumanized more strongly on the HU dimension, the dimension which distinguish humans from nonhuman animals. These findings further support a two-dimensional model of mind perception in faces.

Along with findings from the first two studies, these findings provide further evidence supporting faces as the perceptual basis of mind perception. In particular, faces provide the perceptual cues not only allowing discriminating faces based on whether or not they possess minds in general, but also contributing to nuanced awareness of minds along two distinct dimensions.

Study 4

Since the “new look” movement (Bruner, 1957), psychologists have debated over whether social and cognitive processes influence perceptual ones, the so-called “top-down” effect (for a review, see Firestone & Scholl, 2016).

Researchers have demonstrated similar effects in the perception of face animacy. For example, Fincher and Tetlock (2016) found that faces of norm violators elicit lesser face-typical processing (i.e., configural face processing), which is linked to the denial of minds to faces (Deska et al., 2017; Hugenberg et al., 2015). More direct evidence for the top-down effect entails the findings that participants’ need for social connection facilitates attributing minds to faces (Powers et al., 2014), whereas satisfied social connection with close others renders participants to dehumanize socially distant others (Waytz & Epley, 2012). In addition to the perceivers’ disposition for mind attribution (Epley, Waytz, & Cacioppo, 2007), group identity influences the threshold by which people perceive minds in faces (Hackel et al., 2014; Swiderska et al., 2013). For example, Hackel et al. (2014) found that that participants tended to adopt a higher threshold for perceiving outgroup faces as alive or possessing minds compared with ingroup faces.

Nevertheless, in these studies, the researchers did not manipulate directly the degree of minds the target faces possess but relied on either participants’ tendency for mind attribution or intergroup relations to indirectly influence the extent to which participants might attribute minds to the target faces. Therefore, it remains unknown whether beliefs about others’ minds would directly influence participants’ perception of the categorical boundary of face animacy (Looser & Wheatley, 2010).

In Study 4, instead of manipulating either group membership or a need to connect, I directly manipulated the perceived humanness of faces by using vignettes, which described the protagonists as someone who suffered from a car accident and lost part of his/her mental capacities either to feel (Loss of Experience) or to act (Loss of Agency), or someone who fully recovered from the accident (Complete Human). I then asked participants to judge the perceived categorical boundary of face animacy along a morph continuum created from the protagonist’s face and its well-matched mannequin face. The prediction was that if knowing the protagonist’s status as possessing lesser minds exerts a top-down influence on the perception of its face animacy, in both experimental conditions (Loss of Experience and Loss of Agency), participants should perceive the protagonist’s face to be less human, therefore lowering the threshold compared with the control condition (Complete Human), in which the protagonist was described as fully recovered. Furthermore, I examined whether losing either experience or agency would have differential top-down effect on the threshold judgment of face animacy. Given that human experience is shown to be fundamentally lacking in machines (K. Gray & Wegner, 2012), I predicted that the loss of experience would yield a stronger effect than the loss of agency.

Methods

Participants

I conducted a power analysis using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) for a one-way repeated-measures analysis of variance (ANOVA), assuming a medium effect (η_p^2 = .05), a significance level of .05, and a .5 correlation between measures. This suggested a target *N* of 43 for 95% power in each study. Fifty undergraduates of Emory University (*Mage* = 18.32, *SD* = 2.96, 9 Males) participated in this study. Participants received course credit in compensation. Six participants were removed from analysis due to failure to follow the instructions.

Stimuli

Morph continua. Image morphs of ten unique face identities were selected from Looser and Wheatley (2010). Face morphs were created by morphing each artificial face (100% inanimate) with its well-matched human face (100% animate) using FantaMorph software (Version 5; Abrosoft Co., Beijing, China), resulting in a set of image morphs including the endpoints. Image morphs were presented in a web slider, which allows participants to slide freely through each morph continuum in 1% increments (101 images per morph).

Vignettes. Three vignettes were adapted from previous research (K. Gray, Knickman, & Wegner, 2011; K. Gray & Wegner, 2012). These vignettes described hypothetical protagonists as people who suffered from a car accident and were sent to hospital with one of three possible outcomes: (a) the protagonist lost his/her ability to feel and experience (Loss of Experience), (b) the protagonist lost his/her ability to plan and act (Loss of Agency), and (c) the protagonist fully recovered (see Appendix D for Supplementary materials).

Design and Procedure

The study consisted of three blocks each comprising 10 morph continua. In one block, the protagonists were described as someone who suffered from a car accident and lost the capacity to plan (Loss of Agency). In another block, the same 10 protagonists were described as losing the capacity to feel (Loss of Experience), and yet in another block, they were described as fully recovered (Complete Human). The same vignette was used for all 10 protagonists in each block and the three blocks were presented in a randomized order.

Within each block, participants first read the vignette about each protagonist, and then completed a short questionnaire about his/her perceived level of agency and experience. Finally, participants saw a morph continuum created from the protagonist's face and the face of a well-matched mannequin. Participants were instructed to use a morph slider to indicate the perceived categorical boundary of face animacy along each morph continuum by locating the morph image that just turns alive. The experiment was presented in the web browser by using jsPsych (de Leeuw, 2015).

Planned Data Analysis

The morph percentages across the 10 morph continua were aggregated for each participant and separately for each block as a measure of the threshold of animacy perception. A one-way repeated-measures analysis of variance (ANOVA) was performed on the threshold of animacy with condition as the within-subject factor.

Results and Discussion

One-way repeated-measures ANOVA yielded a significant main effect of condition, $F(2, 86) = 5.65$, $p = .005$, $\eta_p^2 = .12$. As predicted, the threshold of face animacy in the Control condition (Complete Human, $M = 82.01$) was significantly higher than those in the Loss of Experience ($M = 79.42$), $t(43) = 2.96$, $p = .005$, and the Loss of Agency conditions ($M = 80.03$), $t(43) = 2.29$, $p = .027$, whereas there was no statistically significant difference between the two experimental conditions. Results are summarized in Figure 4.

In Study 4, I found that beliefs about the degree of mind someone possesses influences the perception of face animacy. In particular, a person believed to be lacking either experience or agency is perceived as less animate. In other words, more humanlike facial features are needed to judge the face as alive or possessing a mind than if the person had not been dehumanized. Therefore, the current findings corroborate the top-down effect in face animacy perception. Importantly, they demonstrate that the amount of mind ascribed to others can directly alter our interpretation of visual cues indicative of a mind behind the face without involving group membership or a motivation to connect.

Although the findings do not support a distinction between two dimensions of humanness—experience and agency, the statistically nonsignificant results do not preclude the possibility that the difference exists. It is possible that either the vignettes might not generate distinct priming effect or that the measurement (i.e., morph percentage) might not detect the differential effect due to a failure to capture the multidimensional nature of face animacy as did the implicit social cognition measures (e.g., IAT).

Furthermore, the findings are consistent with those by Looser and Wheatley (2010), who demonstrate that as morph percentage increases, faces' perceived animacy as well as their perceived capacities to feel pain or to plan follow a similar cumulative normal function, which are not dissociable by their respective point of subjective equality (PSE). Therefore, the null effect might be best accounted for by the particular measure of categorical boundary used in this experiment. In future research, the differential top-down effect should be further investigated by using alternative paradigms, such as the categorical perception method (e.g., see Looser & Wheatley, 2010, Experiment 2).

Finally, one might argue that findings in the current study are inconsistent with previous evidence suggesting that dehumanization tendencies increase participants' threshold for detecting minds in faces for outgroup members (Hackel et al., 2014; Experiment 1 and 2). Nevertheless, dehumanization might not always lead to an increase in threshold for mind perception. For example, when a particular outgroup is perceived as threatening, participants are inclined to lower the threshold for mind perception. In light of these inconsistent findings, a fruitful direction for future research should further examine how dehumanization influences the threshold for perceiving minds in faces.

General Discussion

Human social interaction depends on understanding other minds. Research on how we gain insights into the minds of others can be broadly organized into three questions: (a) Do they possess minds? (b) What are their states of mind? and (c) Which mental states are responsible for their behavior? (Epley & Waytz, 2010). Researchers argue that understanding the first question of mind awareness is ostensibly the prerequisite for answering the other two (Epley & Waytz, 2010; but see Varga, 2017). Indeed, successful social interactions rely on perceiving others as minded (Epley & Waytz, 2010), and the failure to do so can have dire consequences for both interpersonal and intergroup relations (Haslam & Loughnan, 2014; Zaki, 2014). For example, whether we perceive minds in others determines how we respond to their sufferings: Viewing victims as minded allows us to empathize, that is, to feel in accordance with their affective states, whereas denying their capacity for experience renders us indifferent to and even feel pleased about their misfortune (Cikara, Bruneau, & Saxe, 2011). Denying minds to people excludes them from our circle of moral concerns (K. Gray, Young, &

Waytz, 2012), reduces prosocial behavior; in contrast, it increases antisocial ones toward the dehumanized others (for a review, see Haslam & Loughnan, 2014).

On the other hand, ascribing minds to nonhuman agents (i.e., anthropomorphism; Epley et al., 2007) to some extent creates positive effects, facilitating learning with digital materials (Schneider, Nebel, Beege, & Rey, 2018), eliciting human-oriented interactions (Spexard, Hanheide, & Sagerer, 2007), reducing socially undesirable behavior (Kiesler, Powers, Fussell, & Torrey, 2008) and facilitating social interactions with robots (Duffy, 2003).

Perceiving other minds: What is at stake?

What is missing in the discussion on mind perception is the question of whether our mind awareness can have a genuine perceptual basis, in other words, whether we *perceive* faces as minded with potentials to feel, plan, and interact with our own. The answer to this question should shed light on not only the *problem of other minds* in philosophy but also the uncanny valley phenomenon in robotics.

Although a growing body of literature, including cognitive and social psychology, neuroscience, and human-robot interactions, has begun to investigate these questions (for a review, see Deska & Hugenberg, 2017), the existing literature on face animacy perception is plagued by theoretical and methodological issues surrounding the conception and measurement of minds in faces. In particular, opposing approaches, which I refer to as Mind Addition and Mind Subtraction, raise fundamental challenges for researchers to define what constitutes the perception of minds in faces: Do we perceive different kinds of minds (Dennett, 1996), a hierarchy of minds (Boyer, 1996), or dimensions of minds (H. M. Gray et al., 2007; Haslam, 2006)?

Because answering the epistemological question of whether we have perceptual access to other minds hinges on addressing the ontological question of what minds exist, in the first two studies of my dissertation, I focused on elucidating the dimensions of mind perception in faces. I asked:

Question 1: Do we *perceive* minds in faces?

Question 2: What does the perception of minds in faces entail?

Summary of Study 1 and 2

To address these questions, in Study 1 and 2, I borrowed the implicit association test (IAT) paradigm from research on infrahumanization to examine whether participants would "dehumanize" artificial faces by implicitly associating them more strongly with animal words relative to human words (i.e., animalistic dehumanization). During the experiments, participants were asked to categorize faces and words into respective target (i.e., human vs. artificial faces) and attribute categories (i.e., human vs. animal words) by using designated keys. In some of the trials, the same keys were assigned to both human (artificial) faces and human (animal) words (i.e., congruent trials), whereas in others, the keys for attribute (i.e., word) types were reversed, leaving human (artificial) faces and animal (human) words assigned to the same keys (i.e., incongruent trials). By demonstrating that there is a stronger association, indexed by higher accuracy and shorter reaction time in participants' responses, between human (artificial) faces and human (animal) words relative to the opposite target-attribute association, in Study 1, I established that humans tend to dehumanize artificial faces on the Human Uniqueness dimension, a dimension which distinguishes humans from other animals. In Study 2, I further demonstrated that humans also dehumanize artificial faces on the Human Nature dimension, a dimension which distinguishes humans from machines. In particular, artificial faces were dehumanized more on the Human Nature dimension, providing initial evidence for two-dimensional structure of mind perception in faces.

Research on mind perception in faces is constrained by the operational definition of mind perception in terms of the contrast between human and artificial faces (e.g., Balas & Koldewyn, 2013; Wheatley et al., 2011), a contrast which is akin to Human Nature (i.e., the distinction between humans and machines) but not Human Uniqueness (i.e., the distinction between humans and other animals). Therefore, it would be misleading or incomplete, to say the least, to examine the perception of minds in faces by only attending to one dimension. Given that previous research routinely treat the animacy and species of faces as two distinct constructs (Balas & Auen, 2019; Balas & Koldewyn, 2013; Looser et al., 2013), in Study 3, I sought to extend the conception of mind perception in faces by examining the contrast between human and animal faces. To do this, I asked:

Question 3: Do we dehumanize nonhuman animal faces, and if so, on which dimensions?

Summary of Study 3

In Study 3, I followed the same rationale and methods of the first two studies to examine the extent to which humans might dehumanize nonhuman animal faces. In particular, given that Human Nature is defined by the distinction between humans and machines whereas Human Uniqueness is defined by the distinction between humans and other animals, I predicted that nonhuman animal (e.g., dog) faces should elicit stronger animalistic dehumanization relative to mechanistic dehumanization, whereas artificial faces should elicit the opposite pattern. As predicted, dog faces were dehumanized on both dimensions, more so on the Human Uniqueness dimension, whereas artificial faces were dehumanized more on the Human Nature dimension. Along with findings from the first two studies, these findings provide further support for a two-dimensional structure of mind perception in faces.

Although these three studies showed that artificial and dog faces elicited specific patterns of dehumanization along the human uniqueness and human nature dimensions, one might still wonder whether these findings demonstrate the perception of *mind* in faces. One might argue that instead of perceiving mind in faces, participants may simply associate surface characteristics of the faces (e.g., humanlike facial features) differentially with human and nonhuman words. Nevertheless, this account would not explain why participants demonstrated differential dehumanization given the same surface characteristics of the faces across the two IAT conditions (i.e., animal and automata words).

Nevertheless, there are limitations associated with the IAT paradigm, because this paradigm is designed to examine the relative strength of association between concepts, which cannot be decomposed into separate assessments of single concepts (Nosek et al., 2005). Therefore, findings from Study 1, for example, are consistent with either attributing more humanness to human faces (i.e., humanizing human faces) or attributing less humanness to artificial faces (i.e., dehumanizing artificial faces). To disentangle these two alternative explanations, future research should adopt a GO/NO-GO Association Task (GNAT; Nosek & Banaji, 2001) to achieve separate assessments of artificial and human faces. In addition, in order to achieve higher efficacy and interpretable results using the IAT paradigm, the developers of this paradigm recommend that stimuli should be representative of their categories (e.g., the faces should be unambiguously either animate or inanimate). This requirement, however, prevents the current studies from probing the effect of perceptual similarities between human and nonhuman faces on the magnitude of dehumanization in a nuanced manner (e.g., how would a morphed image of 70% human face and another morphed image of 80% human face differ in the perceptual experience of mindedness they elicit?). To remedy this, future research should consider using the semantic priming paradigm with pictures (Dell'Acqua & Grainger, 1999) to examine, for instance, the extent to which the brief presentation of a face of any given level of human-likeness would facilitate the detection of subsequent stimuli, such as human as opposed to animal words.

In addition to justifying the perception of *minds* by validating the two-dimension structure of mind perception in faces, we can also accomplish this by showing that factors contributing to the attribution of minds to humans or nonhuman agents alter our interpretation of visual cues indicative of minds in faces. To demonstrate this approach, I asked:

Question 4: Is mind perception in faces cognitively penetrable by prior knowledge about others' mental capacities?

Summary of Study 4

Previous research indicates top-down effects of mind perception by showing that participants adopt a higher threshold for perceiving minds in faces of outgroup compared with ingroup members (Hackel et al., 2014) and that individuals with a stronger motivation to connect with others adopt a lower threshold for perceiving minds in faces

(Powers et al., 2014). Overall, these findings suggest that contextual and motivational factors influencing the propensity for mind attribution also affect the interpretation of visual cues indicative of the presence or absence of minds in faces (Epley et al., 2007). Although these findings supposedly suggest the cognitive penetrability of mind perception in faces, I argue that the evidence is inadequate, because without obtaining participants' self-report measures of minds attributed to ingroup vs. outgroup members, researchers can only infer that intergroup dehumanization occurs. Therefore, the existing evidence fails to establish that top-down effects operate via the mediation of intergroup dehumanization.

To remedy this, in Study 4, I directly manipulated participants' prior knowledge about the protagonists' mental capacities using a within-subject design to examine how this cognitive factor would influence their perception of minds in faces. In the experimental conditions, the protagonists were described as someone who had a car accident and were left with brain injuries causing a loss of mental capacities to either formulate plans (e.g., agency) or to feel pain (e.g., experience), whereas in the control condition, the same protagonists were introduced as fully recovered from brain injuries and maintaining intact mental capacities for both agency and experience. The prediction was that knowing the protagonist's lack of either agency or experience should cause participants to judge their faces as possessing lesser minds, therefore shifting the threshold of mind perception toward the doll endpoint along the morph continuum. As predicted, I found that compared with the control condition, participants lowered the threshold for judging faces as "alive," if the participants were told that the protagonists lost certain mental capacities. These findings provide further evidence for the cognitive penetrability of mind perception in faces.

Nevertheless, contrary to my prediction, the findings do not support a differential effect of top-down influence on mind perception in faces between the two dimensions—experience and agency: In fact, lacking either agency or experience similarly rendered the faces of the protagonists to be perceived as possessing lesser minds. The null effect of differential top-down influence might be due to the vignettes' limited ability to elicit dehumanization tendency on distinct dimensions or to the fact that the assessment tool (i.e., morph percentage) might not be sensitive enough to detect any noticeable changes in percepts by top-down influences. Given these limitations, in future research, the shifting perception of minds in faces should be investigated by using alternative paradigms, such as psychophysics to establish the threshold for perceiving minds in faces. Future research may also benefit from using well-documented top-down effects, such as those involving intergroup contexts (e.g., gender and race). Along with the measurement of dehumanization tendencies, researchers will be in a better position to establish the top-down effect via the mediation of mind attribution. Furthermore, researchers may rely on dehumanization of particular targets to provide further insights into nuanced top-down effect on the threshold of mind perception in faces by different types of dehumanization. For example, the dehumanization of sex offenders elicits stronger animalistic dehumanization (Viki, Fullerton, Raggett, Tait, & Wiltshire, 2012), whereas the dehumanization of businesspeople elicits stronger mechanistic dehumanization (Loughnan & Haslam, 2007).

To summarize, in four studies, my dissertation provides preliminary evidence for our perceptual knowledge of other minds via face perception. In particular, the present findings suggest that through faces, we can directly perceive minds in others—humans and nonhuman agents alike, supporting Duddington and Wittgenstein's idea. In addition, these findings are consistent with the phenomenological conception of perceptual processes as providing immediate, rich, information about the particular object that a person perceives. The question of whether face perception is "smart" enough to inform an agent's mental states, however, awaits future investigation (Gallagher, 2008; Varga, 2017).

Based on neural evidence in research on face animacy perception, researchers propose a dual-stage model of face processing (Looser et al., 2013; Wheatley et al., 2011). The theory posits that humans first detect a face (as opposed to a nonface object), and then perceive a mind attached to the face. Implicitly, this theory assumes a distinction between face perception and mind perception, such that the perception of minds is independent from and temporally follows the detection of a face.

The present findings, however, reject this assumption; instead, they suggest that faces, regardless of whether they belong to humans, dogs, or robots, automatically elicit perceptual experience of minds along two distinct dimensions, which define humanity: Human Uniqueness and Human Nature. Whereas Human Uniqueness separates humans from other animals, Human Nature distinguishes humans from machines. In addition, the findings support the continuous conception of minds in faces, following the social psychology tradition in mind perception.

This “continuity” view of minds, however, raises the question of why, if we perceive minds continuously, we have no problem discriminating faces based on whether they appear to possess minds or not (Farid & Bravo, 2012). To address this question, I first propose to investigate the temporal dynamics of mind perception in faces in relation to the uncanny valley phenomenon in robotics. Next, I propose to examine its developmental origins by asking which conception of minds—continuity or discontinuity—emerges first in development.

Temporal Dynamics of Mind Perception in Faces and the Uncanny Valley

Robots are playing an increasingly prominent role in different realms of the modern society, including industry, health care, and education (Broadbent, 2017). Nevertheless, androids, robots with incredible resemblance to humans in appearance and movement, routinely fail to fool humans into mistake them for real people (Farid & Bravo, 2012), but elicit eerie feelings, which hinder their interactions with humans— a phenomenon known as the *uncanny valley* (Mori, MacDorman, & Kageki, 2012).

The uncanny valley poses a puzzle for psychologists, because it points to the ambivalent role humanlike features, such as a face, play in attributing minds to robots and the implications of mind attribution for human-robot interactions. On the one hand, humanlike features facilitate social engagement between humans and machines (Schneider et al., 2018) and elicit increasing human-directed responses in humans (Kiesler et al., 2008). On the other hand, K. Gray and Wegner (2012) argue that humanlike features also elicit the attribution of mind, which is responsible for creating the uncanny valley. In particular, the researchers propose that when robots too closely mimic humans, their humanlike appearance prompts humans to attribute mental capacities, particularly, human experience, to robots. Because such anthropomorphic projections violate humans' prior conceptions of robots and humans as belonging to two distinct ontological categories (i.e., animate and inanimate), they elicit a violation of expectation, which in turn elicits eerie feelings.

Nevertheless, the researchers do not address why the ascription of minds would elicit eerie feelings only for highly humanlike robots (Mathur & Reichling, 2016; Wang & Rochat, 2017), but not for pets, toys, or cartoon characters, which also elicit anthropomorphic projections (Misselhorn, 2009). I argue that at the center of this controversy lies the issue of the temporal dynamics of mind perception (Wheatley et al., 2011). In particular, in relation to faces, I hypothesize that the process of dehumanization, characterized by a perceived decrease of minds in android faces, as opposed to human or mechanical-looking robot faces, as a function of exposure time, provides a more plausible account of the uncanny valley than does the attribution of minds to android faces (Wang, Lilienfeld, & Rochat, 2015).

Although both the dehumanization and the mind perception hypotheses suggest the crucial role of perceiving minds in faces as the triggering the uncanny valley, they differ profoundly in their underlying assumptions regarding how mind perception in faces unfolds over time. In particular, the *Mind Perception Hypothesis* is based on the "mindless until proven minded" assumption that underlies the Mind Addition approach, which predicts a perceived increase of minds in android faces over time. In contrast, the *Dehumanization Hypothesis* is based on the "minded until proven mindless" assumption (Guthrie, 1993) that underlies the Mind Subtraction approach, which predicts a perceived decrease of minds in android faces over time. In particular, I predicted the decrease to occur following 400ms after stimulus onset, the time at which the ERP component of LPP differentiates between human and doll faces suggesting the onset of mind discrimination (Wheatley et al., 2011).

Which of these two scenarios is more likely to describe the temporal dynamics by which we perceive minds in faces? To test this, I presented faces of humans, androids, and mechanical-looking robots at three levels of exposure time ranging from 100ms, 500ms, to 1000ms. Immediately after a brief presentation of the face, participants were asked to rate its perceived animacy (i.e., how alive does it look?) on a 6-point Likert-type scale (e.g., 1 = not alive, 6 = alive). According to the dehumanization hypothesis, I predicted that as a function of exposure time, perceived animacy would undergo a marked decrease for android faces but remain consistently high for human faces and consistently low for mechanical-looking robot faces (see Figure 5). The findings were consistent with these predictions, supporting the dehumanization hypothesis against the mind perception hypothesis. These findings suggest that it is the denial of humanness to android faces rather than anthropomorphic projections that seem to contribute to android faces' perceived eeriness.

In addition to supporting the dehumanization hypothesis against the mind perception hypothesis of the uncanny valley, this study provides further insights into the ways in which we perceive minds in faces. First, at 100ms, both human faces and nonhuman faces, including both android and mechanical-looking robot faces, received ratings of perceived animacy significantly greater than zero. This finding suggests that we do not seem to build our knowledge about minds in faces from ground up, as would be predicted by the Mind Addition account (e.g., the assumption that "a face is mindless until proven minded"). It seems that at first sight, we already have an immediate awareness of how much minds faces possess, depending on their degree of human-likeness. Second, the perceived decrease of minds in android faces suggests that our first impression might not always be accurate such that humans tend to initially perceive more minds in faces than the faces deserve, particularly when nonhuman faces look incredibly humanlike. This initial overattributing tendency is consistent with Guthrie's (1993) theory of anthropomorphism, which posits that anthropomorphic projections impose a complex organization on the perceived objects. Third, although the findings seem to support the Mind Subtraction account, which predicts a decrease in perceived animacy of android faces over time, they also reject the assumption held by the Mind Subtraction account that mind subtraction occurs following an all-or-none law, given that a face either possesses a mind or it does not. Contrary to this assumption, the finding that ratings of animacy of android faces at 500ms and 1000ms did not drop to a comparable level as those of mechanical-looking robot faces is consistent with this interpretation. Finally, the changes in ratings of perceived animacy of android faces suggest that the perception of minds in faces entails accumulating evidence over time and is not completed at once.

Taken together, I argue that the findings imply that the perception of minds in faces entails at least two processes, in particular, both the attribution and the discrimination of minds to faces. Whereas my dissertation demonstrates that mind attribution is multidimensional and *continuous*, and belongs to a perceptual process, the findings from the current research on the uncanny valley suggest that mind discrimination is *discontinuous* (e.g., dichotomous) and belongs to a decision-making process. Mind discrimination is discontinuous and belongs to cognitive processes, particularly because the uncanny valley points to the ambiguous roles faces play: Faces elicit perception that indicates both the presence and the absence of minds. It is the need to transform this mixed package of information into a definite categorization of the face as either minded or not that requires cognitive processes, such as decision-making, to supplement perception to allow humans to accurately discriminate minds in faces to navigate the social world. Mind attribution of faces allows people to cast a wider net for seeking possible social connections (Powers et al., 2014), but it is subject to false alarms, such as overattributing minds to mindless faces. In contrast, mind discrimination of faces reduces false alarms and allows people to allocate limited social-cognitive resources more efficiently to faces worthy of thoughts, feelings, and conversations.

To summarize, the uncanny valley lies at the frontier of research on the perceptual basis of mind perception, particularly in faces. Therefore, I argue that the apparent contradiction between the "continuity" and "discontinuity" views of minds, instead of posing a challenge for perceiving minds in faces, brings benefits for both maximizing the opportunities for establishing new social relationships and increasing mind attribution accuracy.

Origins of Mind Perception and Anthropomorphism

Consistent with the Mind Addition approach is the idea in social cognition research that humans understand nonhuman entities in similar ways (e.g., in terms of mental state attribution) as they understand each other. This continuity hypothesis therefore calls for the integration between different research traditions, such as between anthropomorphism and dehumanization and between interpersonal and human-robot interactions (Fiske et al., 2007; H. M. Gray et al., 2007; Haslam, 2006). In contrast, there is also a discontinuity hypothesis that prevails the cognitive and developmental literature (Opfer & Gelman, 2011), which posits that humans and nonhuman entities represent two ontological categories, whereby domain-specific principles prevent perceiving anything in between (Boyer, 1996; Farah & Heberlein, 2007). This discontinuity hypothesis underlies the Mind Subtraction approach. The question is: Are these two opposing views reconcilable when we perceive minds in faces? In addition to addressing this question by pointing to the temporal dynamics of mind perception in faces, I argue that researchers can also draw inspiration from research on the development of face processing during infancy.

Before delving into the developmental research on infant face processing, it is crucial to link mind discrimination to the closely related concept of person-object distinction, which constitutes a basic ontological distinction and marks infants' emergence as a social being (Legerstee, 1992). But when do humans begin to make this distinction? According to Piaget (1954), the distinction between physical and social world does not emerge until two years of age. Nevertheless, numerous studies have shown that young infants begin to make ontological distinctions between such domains as animate and inanimate few months after birth (for a review, see Opfer & Gelman, 2011). Boyer (1996) argues that it is particularly due to this early conceptual development that infants cannot project humanlike characteristics onto nonhuman entities (i.e., anthropomorphism) without violating domain-specific expectations for humans and nonhuman beings. Therefore, he argues that anthropomorphic projections common to religious beliefs are counterintuitive and should be accounted for by cultural transmission of religious assumptions.

Here, I argue that the domain-specific knowledge for what humans are is learned rather than innate; in addition, our knowledge of what distinguish humans from objects can originate from motion (e.g., self-propelled motion) and contingency cues as much as the degree of physical similarities to humans. Therefore, anthropomorphism based on human physical features, unlike most anthropomorphism in religious systems, which does not specify supernatural agents' physical resemblance to people, might have a more intuitive, perceptual basis. In particular, I argue that the development of face processing in infants can provide useful insights into the developmental roots of how humans have learned to associate faces with minds intricately (e.g., we recognize that faces can suggest both the presence and absence of minds). In what follows, I briefly review the relevant literature, based on which I speculate the potential role of mind perception in faces plays in the development of face processing in infants.

Newborn infants preferentially orient toward faces and facelike patterns (Goren, Sarty, & Wu, 1975). Various theories have been proposed to account for this early face preference. A sensory hypothesis (Banks & Salapatek, 1981) suggests that infants are predisposed to attend to visual stimuli with high "stimulus energy" as measured by stimuli's amplitude spectrum using Fourier transformations. In contrast, a social hypothesis (Gibson, 1969) proposes that infants are predisposed to attend to social stimuli, such as facelike configurations due to the familiarity and social significance of faces. To examine which hypothesis better accounts for face preference at birth and at 2 months, Kleiner and Banks (Kleiner, 1987; Kleiner & Banks, 1987) employed an amplitude—phase spectra hybrid technique (Piotrowski & Campbell, 1982), which combined the stimulus energy information of one stimulus with the identity information of another stimulus (e.g., a visual stimuli combining a face's amplitude spectrum and a lattice's phase spectrum). The researchers predicted that according to the sensory hypothesis, infants' face preference would be driven by the amplitude spectrum of a face (high stimulus energy), regardless of its phase spectrum (social significance), whereas according to the social hypothesis, it would predict the opposite pattern. What Kleiner and her colleague found was that images' amplitude spectrum predicted neonates' looking preference, regardless of their phase spectrum (supporting the sensory hypothesis), whereas the images' phase but not amplitude spectrum predicted 2-month-olds' looking preference (supporting the social hypothesis).

Research has also proposed a "top-heavy" bias to explain newborns' face preference, which posits that infants possess a visual bias that favors configuration consisting of more elements in the upper half than in the lower half of a visual stimulus (Cassia, Turati, & Simion, 2004; Nelson, 2001; Simion, Cassia, Turati, & Valenza, 2001; Turati, Simion, Milani, & Umiltà, 2002). Studies showed that newborns looked equally long at faces (e.g., schematic and veridical face images) paired with an equally complex top-heavy nonface configuration (Cassia et al., 2004; Turati et al., 2002). In contrast, 3 month old infants' looking preference is not attributable to a "top-heavy" bias, suggesting that compared with newborns, 3-month-old infants' visual preference is better explained by a face-specific, social, rather than by a domain-general (e.g., a top-heavy bias), asocial, principle (Simion, Leo, Turati, Valenza, & Barba, 2007).

Finally, 5-month-olds demonstrated preferential looking toward schematic face-like patterns only if the internal features of the patterns moved, whereas movements of internal facial features did not elicit preferential looking in younger infants (Johnson, Dziurawiec, Bartrip, & Morton, 1992). Furthermore, 7 to 8 month but not 5 months old infants demonstrated a preference for vertical face-like movements (e.g., opening-closing movements of the eyes and the mouth) compared with horizontal non face-like movements (Ichikawa, Kanazawa, & Yamaguchi, 2011). These findings suggest that veridical facial movements play a more important role in governing the looking preference for older than for younger infants. In particular, older infants demonstrate the ability to consider veridical face-like movements into the perception of faces.

Taken together, these findings suggest that with age, infant face preference is increasingly governed by face-specific mechanisms, which highlight the necessity of humanlike features in the facial stimuli for attracting and sustaining infants' attention (Chien, 2011; Gliga, Elsabbagh, Andravizou, & Johnson, 2009; Ichikawa et al., 2011; Ichikawa, Tsuruhara, Kanazawa, & Yamaguchi, 2013; Turati, Valenza, Leo, & Simion, 2005).

In other words, this developmental trajectory highlights infants' growing sensitivity to perceptual cues in conspecific (i.e., human) faces to not only detect a face but also discriminate faces based on their level of human-likeness. Although these studies do not directly demonstrate the role of mind perception in the developmental changes in infant face preference, the evidence is consistent with this hypothesis.

Importantly, the development of face preference in infants does not follow a monotonic trend but undergoes a U-shape trajectory. For example, infants' visual tracking of face-like stimuli undergoes a decline around 4 to 6 weeks, followed by a rebound at 2 months, which is accompanied by the acquirement of more sophisticated face processing abilities (Johnson et al., 1992; Johnson, Farroni, Brockbank, & Simion, 2000; Morton & Johnson, 1991). Based on Horn's filial imprinting model in chicks (Horn & Johnson, 1989; Johnson, Bolhuis, & Horn, 1985), Johnson and colleagues propose a 2-process theory, which posits that infant face preference is governed by 2 distinct mechanisms: CONSPEC and CONLERN. The CONSPEC mechanism is an experience-independent subcortically controlled system, which imposes a predisposition in newborn infants to preferentially orient toward faces and face-like configurations. In contrast, the CONLERN mechanism is an experience-dependent cortically controlled system, which guides the learning of conspecific faces.

Given that CONLERN is an acquired cortical specialization for learning various other aspects of face processing other than face detection, it plays a crucial role in the development of face expertise primarily preserved for own-species and own-race faces own-race (Kelly et al., 2009; Kelly et al., 2007) and own-species (Pascalis, de Haan, & Nelson, 2002; Pascalis et al., 2005), collectively known as the perceptual narrowing effect (Nelson, 2001).

Adult research points to artificial faces as yet another social categories with which we lack face expertise (Balas & Pacella, 2015; Crookes et al., 2015). Because the very types of faces (e.g., other-species, other-race, and computer-generated faces) with which we demonstrate diminished face expertise represent the categories of humans or nonhuman agents in which we attribute lesser minds, studying the face-mind linkage by tracing its developmental roots in the correlations between face expertise and intergroup bias in infants should be a fruitful future direction, whereby the dual-process model would be an ideal starting point. Assuming that researchers can find when infants demonstrate the ability to associate more humanlike faces with more minds (e.g., mind attribution) and when they begin to demonstrate the categorical perception of face animacy (e.g., mind discrimination), these findings can eventually shed light on the continuity vs. discontinuity debate by informing which mode of perceiving minds in faces emerges first in life.

References

Bain, P., Park, J., Kwok, C., & Haslam, N. (2009). Attributing human uniqueness and human nature to cultural groups: Distinct forms of subtle dehumanization. *Group Processes & Intergroup Relations, 12*(6), 789-805. doi: 10.1177/1368430209340415

Balas, B., & Auen, A. (2019). Perceiving animacy in own-and other-species faces. *Frontiers in Psychology, 10*. doi: 10.3389/fpsyg.2019.00029

Balas, B., & Koldewyn, K. (2013). Early visual ERP sensitivity to the species and animacy of faces. *Neuropsychologia, 51*(13), 2876-2881. doi: 10.1016/j.neuropsychologia.2013.09.014

Balas, B., & Pacella, J. (2015). Artificial faces are harder to remember. *Computers in Human Behavior, 52*, 331-337. doi: 10.1016/j.chb.2015.06.018

Banerjee, K., & Bloom, P. (2014). Why did this happen to me? Religious believers' and non-believers' teleological reasoning about life events. *Cognition, 133*(1), 277-303. doi: 10.1016/j.cognition.2014.06.017

Banks, M. S., & Salapatek, P. (1981). Infant pattern vision: A new approach based on the contrast sensitivity function. *Journal of Experimental Child Psychology, 31*(1), 1-45. doi: 10.1016/0022-0965(81)90002-3

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences, 4*(1), 29-34. doi: 10.1016/S1364-6613(99)01419-9

Barrett, J. L., & Keil, F. C. (1996). Conceptualizing a nonnatural entity: Anthropomorphism in god concepts. *Cognitive Psychology, 31*(3), 219-247. doi: 10.1006/cogp.1996.0017

Bilewicz, M., Imhoff, R., & Drogosz, M. (2011). The humanity of what we eat: Conceptions of human uniqueness among vegetarians and omnivores. *European Journal of Social Psychology, 41*(2), 201-209. doi: 10.1002/ejsp.766

Boyer, P. (1996). What makes anthropomorphism natural: Intuitive ontology and cultural representations. *The Journal of Royal Anthropological Institute, 2*(1), 83-97. doi: 10.2307/3034634

Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annu Rev Psychol, 68*(1), 627-652. doi: 10.1146/annurev-psych-010416-043958

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64*(2), 123-152. doi: 10.1037/h0043805

Carpenter, T., Pogacar, R., Pullig, C., Kouril, M., Aguilar, S., LaBouff, J., . . . Chakroff, A. (2018). Survey-based implicit association tests: A methodological and empirical analysis. *PsyArXiv*. doi: 10.31234/osf.io/hgy3z

Cassia, V. M., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science, 15*(6), 379-383. doi: 10.1111/j.0956-7976.2004.00688.x

Chien, S. H.-L. (2011). No more top-heavy bias: Infants and adults prefer upright faces but not top-heavy geometric or face-like patterns. *Journal of Vision, 11*(6), 13, 11-14. doi: 10.1167/11.6.13

Cikara, M., Bruneau, E. G., & Saxe, R. R. (2011). Us and Them: Intergroup Failures of Empathy. *Current Directions in Psychological Science, 20*(3), 149-153. doi: 10.1177/0963721411408713

Crookes, K., Ewing, L., Gildenhuys, J. D., Kloth, N., Hayward, W. G., Oxner, M., . . . Rhodes, G. (2015). How well do computer-generated faces tap face expertise? *PLoS One, 10*(11), e0141353. doi: 10.1371/journal.pone.0141353

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(1), 1-12. doi: 10.3758/s13428-014-0458-y

Dell'Acqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition*, *73*(1), B1-B15. doi: 10.1016/S0010-0277(99)00049-9

Dennett, D. C. (1996). *Kinds of minds: Toward an understanding of consciousness*. New York, NY: Basic Books.

Deska, J. C., Almaraz, S. M., & Hugenberg, K. (2017). Of mannequins and men: Ascriptions of mind in faces are bounded by perceptual and processing similarities to human faces. *Social Psychological and Personality Science, 8*(2), 183-190. doi: 10.1177/1948550616671404

Deska, J. C., & Hugenberg, K. (2017). The face-mind link: Why we see minds behind faces, and how others' minds change how we see their face. *Social and Personality Psychology Compass, 11*(12), e12361. doi: 10.1111/spc3.12361

Duddington, N. A. (1918). Our knowledge of other minds. *Proceedings of the Aristotelian Society, 19*, 147-178.

Duffy, B. R. (2003). Anthropomorphism and the social robot *Robotics and Autonomous Systems, 42*, 177-190. doi: 10.1016/S0921-8890(02)00374-3

Eddy, T. J., Gallup, G. G., & Povinelli, D. J. (1993). Attribution of Cognitive States to Animals: Anthropomorphism in Comparative Perspective. *Journal of Social Issues, 49*(1), 87-101. doi: 10.1111/j.1540-4560.1993.tb00910.x

Epley, N., & Waytz, A. (2010). Mind Perception. In S. T. Fiske, D. T. Gilbert & G. Lindzey (Eds.), *Handbook of Social Psychology* (Vol. 1, pp. 498-541). Hoboken: NJ: John Wiley.

Epley, N., Waytz, A., & Cacioppo, J. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864-886. doi: 10.1037/0033-295X.114.4.864

Farah, M. J., & Heberlein, A. S. (2007). Personhood and neuroscience: naturalizing or nihilating? *Am J Bioeth, 7*(1), 37-48. doi: 10.1080/15265160601064199

Farid, H., & Bravo, M. J. (2012). Perceptual discrimination of computer generated and photographic faces. *Digital Investigation, 8*(3–4), 226-235. doi: 10.1016/j.diin.2011.06.003

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi: 10.3758/BF03193146

Fincher, K. M., & Tetlock, P. E. (2016). Perceptual dehumanization of faces is activated by norm violations and facilitates norm enforcement. *Journal of Experimental Psychology-General, 145*(2), 131-146. doi: 10.1037/xge0000132

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences, 39*, e229. doi: 10.1017/S0140525X15000965

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77-83. doi: 10.1016/j.tics.2006.11.005

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1431), 459-473. doi: 10.1098/rstb.2002.1218

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, *17*(2), 535-543. doi: 10.1016/j.concog.2008.03.003

Gallagher, S., & Varga, S. (2014). Social constraints on the direct perception of emotions and intentions. *Topoi, 33*(1), 185-199. doi: 10.1007/s11245-013-9203-x

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493-501. doi: 10.1016/S1364-6613(98)01262-5

Gao, T., & Scholl, B. J. (2011). Chasing vs. stalking: Interrupting the perception of animacy. *Journal of Experimental Psychology: Human Perception and Performance, 37*(3), 669-684. doi: 10.1037/a0020735

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. H. (2009). Faces attract infants' attention in complex displays. *Infancy, 14*(5), 550-562. doi: 10.1080/15250000903144199

Goldman, A. (2013). *Joint ventures: mindreading, mirroring, and embodied cognition*. Oxford: Oxford University Press.

Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts and theories*. Cambridge, MA: MIT Press.

Gordon, R. M. (2008). Folk psychology as simulation. In W. G. Lycan & J. J. Prinz (Eds.), *Mind and cognition: An anthology., 3rd ed.* (pp. 369-378). Malden: Blackwell Publishing.

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics, 56*(4), 544-549.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619. doi: 10.1126/science.1134475

Gray, K., Knickman, T. A., & Wegner, D. M. (2011). More dead than dead: Perceptions of persons in the persistent vegetative state. *Cognition, 121*(2), 275-280. doi: 10.1016/j.cognition.2011.06.014

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition, 125*(1), 125-130. doi: 10.1016/j.cognition.2012.06.007

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*(2), 101-124. doi: 10.1080/1047840X.2012.651387

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *J Pers Soc Psychol, 74*(6), 1464-1480. doi: 10.1037/0022-3514.74.6.1464

Grossmann, T. (2017). The eyes as windows into other minds: An integrative perspective. *Perspectives on Psychological Science, 12*(1), 107-121. doi: 10.1177/1745691616654457

Guthrie, S. (1993). *Faces in the clouds*. New York: NY: Oxford University Press.

Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. *Journal of Experimental Social Psychology, 52*, 15-23. doi: 10.1016/j.jesp.2013.12.001

Hackel, L. M., Mende-Siedlecki, P., Looser, C., & Van Bavel, J. J. (2015). Extracting social meaning from a face: The neural substrates and behavioral repercussions of mind perception. *Social Science Research Network*. doi: 10.2139/ssrn.2662842

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10*(3), 252-264. doi: 10.1207/s15327957pspr1003_4

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annu Rev Psychol, 65*, 399-423. doi: 10.1146/annurev-psych-010213-115045

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science, 344*(6190), 1243091. doi: 10.1126/science.1243091

Horn, G., & Johnson, M. H. (1989). Memory systems in the chick: Dissociations and neuronal analysis. *Neuropsychologia, 27*(1), 1-22. doi: 10.1016/0028-3932(89)90086-9

Hugenberg, K., Young, S., Rydell, R. J., Almaraz, S., Stanko, K. A., See, P. E., & Wilson, J. P. (2015). The face of humanity: Configural face processing influences ascriptions of humanness. *Social Psychological and Personality Science*, *7*(2), 167-175. doi: 10.1177/1948550615609734

Hume, D. (1757/1957). *The natural history of religion*: Stanford University Press.

Ichikawa, H., Kanazawa, S., & Yamaguchi, M. K. (2011). The movement of internal facial features elicits 7 to 8-month-old infants' preference for face patterns. *Infant and Child Development, 20*(5), 464-474. doi: 10.1002/icd.724

Ichikawa, H., Tsuruhara, A., Kanazawa, S., & Yamaguchi, M. K. (2013). Two- to three-month-old infants prefer moving face patterns to moving top-heavy patterns. *Japanese Psychological Research, 55*(3), 254-263. doi: 10.1111/j.1468-5884.2012.00540.x

Ishiguro, H. (2006). Android science: conscious and subconscious recognition. *Connection Science, 18*(4), 319-332. doi: 10.1080/09540090600873953

Johnson, M. H., Bolhuis, J. J., & Horn, G. (1985). Interaction between acquired preferences and developing predispositions during imprinting. *Animal Behaviour, 33*(3), 1000-1006. doi: 10.1016/S0003-3472(85)80034-8

Johnson, M. H., Dziurawiec, S., Bartrip, J., & Morton, J. (1992). The effects of movement of internal features on infants' preferences for face-like stimuli. *Infant Behavior and Development, 15*(1), 129-136. doi: 10.1016/0163-6383(92)90011-T

Johnson, M. H., Farroni, T., Brockbank, M., & Simion, F. (2000). Preferential orienting to faces in 4-month-olds: Analysis of temporal–nasal visual field differences. *Developmental Science, 3*(1), 41-45. doi: 10.1111/1467-7687.00097

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: evidence toward universality? *Journal of Experimental Child Psychology, 104*(1), 105-114. doi: 10.1016/j.jecp.2009.01.006

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: evidence of perceptual narrowing. *Psychological Science, 18*(12), 1084-1089. doi: 10.1111/j.1467-9280.2007.02029.x

Kelman, H. C. (1976). Violence without restraint: Reflections on the dehumanization of victims and victimizers. In G. M. Kren & L. H. Rappoport (Eds.), *Varieties of psychohistory* (pp. 282-314). New York: Springer.

Kiesler, S., Powers, A., Fussell, S. R., & Torrey, C. (2008). Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition, 26*(2), 169-181. doi: 10.1521/soco.2008.26.2.169

Kleiner, K. (1987). Amplitude and phase spectra as indices of infants' pattern preferences. *Infant Behavior and Development, 10*(1), 49-59. doi: 10.1016/0163-6383(87)90006-3

Kleiner, K., & Banks, M. S. (1987). Stimulus energy does not account for 2-month-olds' face preferences. *Journal of Experimental Psychology: Human Perception and Performance, 13*(4), 594-600. doi: 10.1037/0096-1523.13.4.594

Kwan, V. S. Y., & Fiske, S. T. (2008). Missing links in social cognition: The continuum from nonhuman agents to dehumanized humans. *Social Cognition*, *26*(2), 125-128. doi: DOI 10.1521/soco.2008.26.2.125

Legerstee, M. (1992). A review of the animate-inanimate distinction in infancy: Implications for models of social and cognitive knowing. *Early Development and Parenting, 1*(2), 59-67. doi: 10.1002/edp.2430010202

Leyens, J.-P., Demoulin, S., Vaes, J., Gaunt, R., & Paladino, M. P. (2007). Infra-humanization: The Wall of Group Differences. *Social Issues and Policy Review, 1*(1), 139-172. doi: 10.1111/j.1751-2409.2007.00006.x

Leyens, J.-P., Rodriguez-Perez, A., Rodriguez-Torres, R., Gaunt, R., Paladino, M. P., Vaes, J., & Demoulin, S. (2001). Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology, 31*(4), 395-411. doi: 10.1002/Ejsp.50

Looser, C. E., Guntupalli, J. S., & Wheatley, T. (2013). Multivoxel patterns in face-sensitive temporal regions reveal an encoding schema based on detecting life in a face. *Social Cognitive Affective Neuroscience, 8*(7), 799-805. doi: 10.1093/scan/nss078

Looser, C. E., & Wheatley, T. (2010). The tipping point of animacy: How, when, and where we perceive life in a face. *Psychological Science, 21*(12), 1854-1862. doi: 10.1177/0956797610388044

Loughnan, S., & Haslam, N. (2007). Animals and androids: implicit associations between social categories and nonhumans. *Psychological Science, 18*(2), 116-121. doi: 10.1111/j.1467-9280.2007.01858.x

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition, 146*, 22-32. doi: 10.1016/j.cognition.2015.09.008

Misselhorn, C. (2009). Empathy with inanimate objects and the uncanny valley. *Minds and Machines, 19*(3), 345-359. doi: 10.1007/s11023-009-9158-2

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *Robotics & Automation Magazine, IEEE, 19*(2), 98-100. doi: 10.1109/MRA.2012.2192811

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review, 98*(2), 164-181. doi: 10.1037/0033-295x.98.2.164

Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development, 10*(1-2), 3-18. doi: 10.1002/icd.239

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*(6), 625-666. doi: 10.1521/soco.19.6.625.20886

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: Ii. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*(2), 166-180. doi: 10.1177/0146167204271418

Opfer, J. E., & Gelman, S. A. (2011). Development of the animate-inanimate distinction. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development, 2nd ed.* (pp. 213-238): Wiley-Blackwell.

Paladino, M.-P., Leyens, J.-P., Rodriguez, R., Rodriguez, A., Gaunt, R., & Demoulin, S. (2002). Differential association of uniquely and non uniquely human emotions with the ingroup and the outgroup. *Group Processes & Intergroup Relations, 5*(2), 105-117. doi: 10.1177/1368430202005002539

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science, 296*(5571), 1321-1323. doi: 10.1126/science.1070223

Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences, 102*(14), 5297-5300. doi: 10.1073/pnas.0406627102

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Piotrowski, L. N., & Campbell, F. W. (1982). A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception, 11*(3), 337-346. doi: 10.1068/p110337

Powers, K. E., Worsham, A. L., Freeman, J. B., Wheatley, T., & Heatherton, T. F. (2014). Social connection modulates perceptions of animacy. *Psychological Science, 25*(10), 1943-1948. doi: 10.1177/0956797614547706

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515-526. doi: 10.1017/S0140525X00076512

Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2004). Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *Journal of Vision, 4*(1), 13-21. doi: 10.1167/4.1.2

Santos, N. S., David, N., Bente, G., & Vogeley, K. (2008). Parametric induction of animacy experience. *Consciousness and Cognition, 17*(2), 425-437. doi: 10.1016/j.concog.2008.03.012

Schneider, S., Nebel, S., Beege, M., & Rey, G. D. (2018). Anthropomorphism in decorative pictures: Benefit or harm for learning? *Journal of Educational Psychology, 110*(2), 218-232. doi: 10.1037/edu0000207

Shannon, R. W., Patrick, C. J., Venables, N. C., & He, S. (2013). ‘Faceness’ and affectivity: Evidence for genetic contributions to distinct components of electrocortical response to human faces. *NeuroImage, 83*, 609-615. doi: 10.1016/j.neuroimage.2013.06.014

Simion, F., Cassia, V. M., Turati, C., & Valenza, E. (2001). The origins of face perception: specific versus non-specific mechanisms. *Infant and Child Development, 10*(1-2), 59-65. doi: 10.1002/icd.247

Simion, F., Leo, I., Turati, C., Valenza, E., & Barba, B. D. (2007). How face specialization emerges in the first months of life. *From Action to Cognition, 164*, 169-185. doi: 10.1016/S0079-6123(07)64009-6

Spexard, T. P., Hanheide, M., & Sagerer, G. (2007). Human-oriented interaction with an anthropomorphic robot. *IEEE Transactions on Robotics, 23*(5), 852-862. doi: 10.1109/TRO.2007.904903

Staub, E. (1989). *The roots of evil: The origins of genocide and other group violence*. New York: Cambridge University Press.

Swiderska, A., Krumhuber, E. G., & Kappas, A. (2013). *No matter how real: Outgroup faces convey less humanness.* Paper presented at the Proceedings of the 7th Workshop on Emotion and Computing - Current Research and Future Impact, Koblenz, Germany.

Turati, C., Simion, F., Milani, I., & Umiltà, C. (2002). Newborns' preference for faces: What is crucial? *Developmental Psychology, 38*(6), 875-882. doi: 10.1037/0012-1649.38.6.875

Turati, C., Valenza, E., Leo, I., & Simion, F. (2005). Three-month-olds' visual preference for faces and its underlying visual processing mechanisms. *Journal of Experimental Child Psychology, 90*(3), 255-273. doi: 10.1016/j.jecp.2004.11.001

Varga, S. (2017). The case for mind perception. *Synthese, 194*(3), 787-807. doi: 10.1007/s11229-015-0994-8

Vidal, D. (2007). Anthropomrophism or sub-anthropomorphism? An anthropological approach to gods and robots. . *Journal of the Royal Anthropological Institute, 13*(4), 917-933. doi: 10.1111/j.1467-9655.2007.00464.x

Viki, G. T., & Abrams, D. (2003). Infra-humanization: Ambivalent sexism and the attribution of primary and secondary emotions to women. *Journal of Experimental Social Psychology, 39*(5), 492-499. doi: 10.1016/S0022-1031(03)00031-3

Viki, G. T., Fullerton, I., Raggett, H., Tait, F., & Wiltshire, S. (2012). The role of dehumanization in attitudes toward the social exclusion and rehabilitation of sex offenders. *Journal of Applied Social Psychology, 42*(10), 2349-2367. doi: 10.1111/j.1559-1816.2012.00944.x

Viki, G. T., Winchester, L., Titshall, L., Chisango, T., Pina, A., & Russell, R. (2006). Beyond secondary emotions: The infrahumanization of outgroups using human-related and animal-related words. *Social Cognition, 24*(6), 753-775. doi: 10.1521/soco.2006.24.6.753

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology, 19*(4), 393-407. doi: 10.1037/gpr0000056

Wang, S., & Rochat, P. (2017). Human perception of animacy in light of the uncanny valley phenomenon. *Perception, 46*(12), 1386-1411. doi: 10.1177/0301006617722742

Waytz, A., & Epley, N. (2012). Social connection enables dehumanization. *Journal of Experimental Social Psychology, 48*(1), 70-76. doi: 10.1016/j.jesp.2011.07.012

Waytz, A., Epley, N., & Cacioppo, J. (2010). Social cognition unbound: Insights into anthropomorphism and dehumanization. *Current Directions in Psychological Science, 19*(1), 58-62. doi: 10.1177/0963721409359302

Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA, US: MIT Press.

Wellman, H. M. (1992). *The child's theory of mind*. Cambridge, MA, US: The MIT Press.

Wheatley, T., Weinberg, A., Looser, C., Moran, T., & Hajcak, G. (2011). Mind perception: real but not artificial faces sustain neural activity beyond the N170/VPP. *PLoS One, 6*(3), e17960. doi: 10.1371/journal.pone.0017960

Zaki, J. (2014). Empathy: A motivated account. *Psychological Bulletin, 140*(6), 1608-1647. doi: 10.1037/a0037679

Table 1

Stimulus Set	**Attribute**	**N**	**Mean**	**SD**	**95% CI**	***t* statistic**	***p***
Sims	Animal	79	0.52	0.34	[0.44, 0.60]	13.57	1.6e-22
Sims	Automata	79	0.63	0.36	[0.55, 0.71]	15.74	3.2e-26
LW	Animal	79	0.58	0.31	[0.50, 0.66]	16.41	2.6e-27
LW	Automata	79	0.64	0.38	[0.56, 0.72]	15.07	4e-25
UV	Animal	79	0.63	0.31	[0.57, 0.69]	18.18	4.4e-30
UV	Automata	79	0.71	0.34	[0.63, 0.79]	18.84	4.6e-31

Table 1. Mean D scores and one-sample *t* test across three stimulus sets (Sims, LW, and UV) and two attribute types (Animal vs. Automata).

Table 2

Stimulus Set	Animal	Automata	*t* statistic	*p*
Sims	0.52	0.63	-2.88	0.0052
LW	0.58	0.64	-1.5	0.14
UV	0.63	0.71	-2.12	0.037

Table 2. D scores by attribute type (Animal vs. Automata) and results of paired-sample t test across three stimulus sets (Sims, LW, and UV). Error bars represent one standard error from the mean D score.

Table 3

Target Face	**Attribute**	**N**	**Mean**	**SD**	**95% CI**	***t* statistic**	***p***
Dog	Animal	43	0.73	0.35	[0.62, 0.83]	13.79	1.7e-17
Dog	Automata	43	0.59	0.3	[0.50, 0.68]	12.77	2.3e-16
Artificial	Animal	43	0.57	0.31	[0.48, 0.66]	12.24	9.8e-16
Artificial	Automata	43	0.68	0.36	[0.58, 0.79]	12.35	7.3e-16

Table 3. Mean D scores and one-sample *t* test for target type (Dog Face vs. Artificial Face) and attribute type (Animal vs. Automata)

Table 4

Target Face	Animal	Automata	*t* statistic	*p*
Dog	0.73	0.59	2.40	0.021
Artificial	0.57	0.68	-1.87	0.069

Table 4. D scores by target type (Dog Face vs. Artificial Face) and attribute type (Animal vs. Automata) and results of paired-sample t test

Figure 1

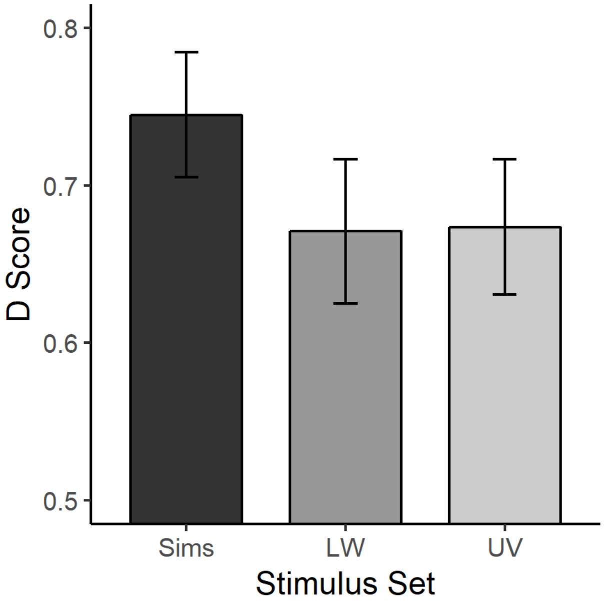


Figure 1. D scores from each of the three IATs corresponding to three stimulus sets: Sims, LW, and UV. A positive D score indicates a stronger association between human (artificial) faces with human (animal) than with animal (human) words. Error bars represent one standard error from the mean D score.

Figure 2

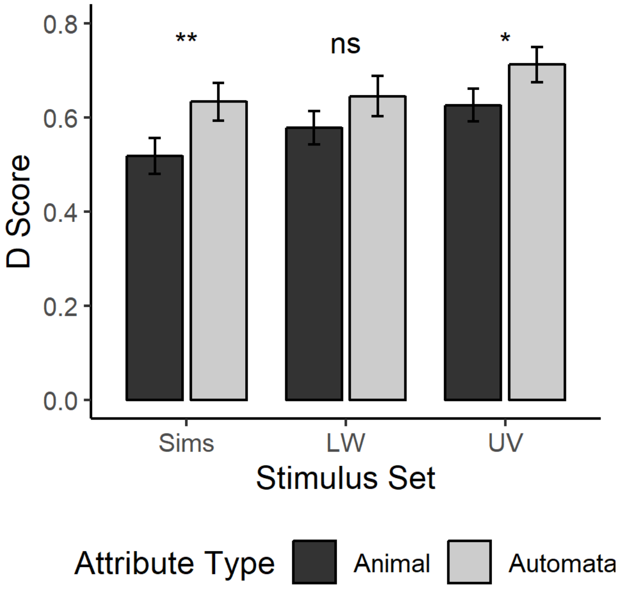


Figure 2. D scores from each of the six IATs corresponding to three stimulus sets (Sims, LW, and UV) and two attribute types (human vs. animal words, human vs. automata words). A positive D score indicates a stronger association between human (artificial) faces with human (animal or automata) than with animal/automata (human) words. Error bars represent one standard error from the mean D score.

*p < .05. **p < .01. ***p < .001.

Figure 3

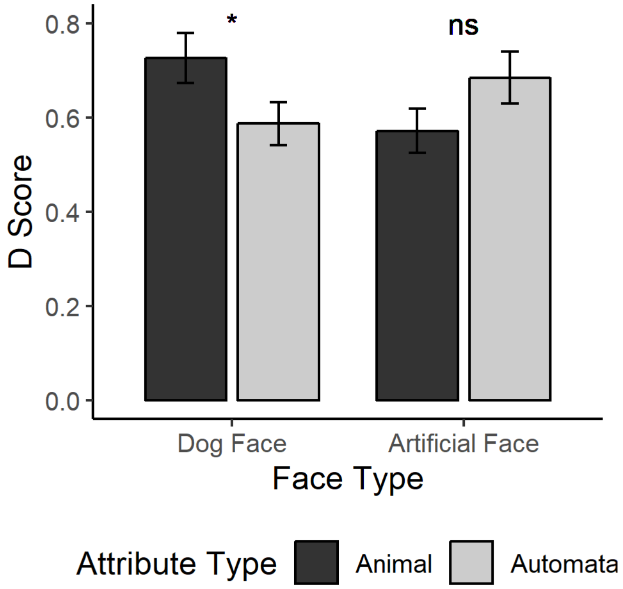


Figure 3. D scores from each of the four IATs corresponding to two target types (Dog Face vs. Artificial Face) and two attribute types (human vs. animal words, human vs. automata words). A positive D score indicates a stronger association between human (artificial/dog) faces with human (animal/automata) than with animal/automata (human) words. Error bars represent one standard error from the mean D score.

**p* < .05. ***p* < .01. ****p* < .001.

Figure 4

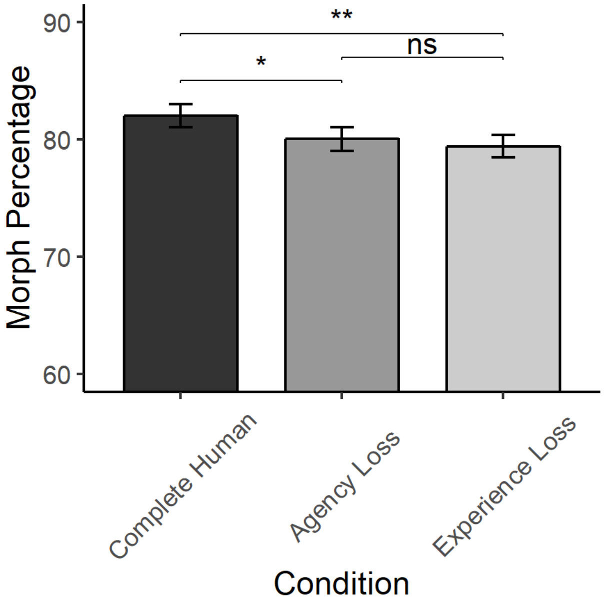


Figure 4. Thresholds (morph percentages) of face animacy when the protagonists were described as fully recovered from the card accident (Complete Human), losing the mental capacity to formulate plans (Loss of Agency), or losing the mental capacity to feel pain (Loss of Experience). A higher threshold indicates that the face needs to be more humanlike to be perceived as being alive or possessing a mind. Error bars represent one standard error from the mean morph percentage.

**p* < .05. ***p* < .01. ****p* < .001.

Figure 5

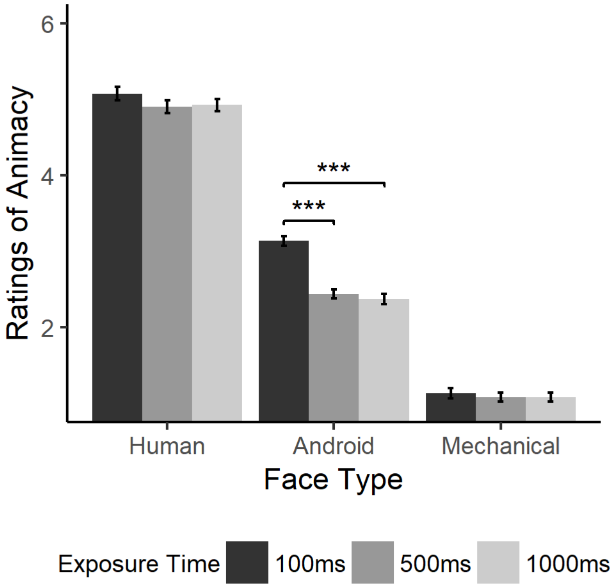
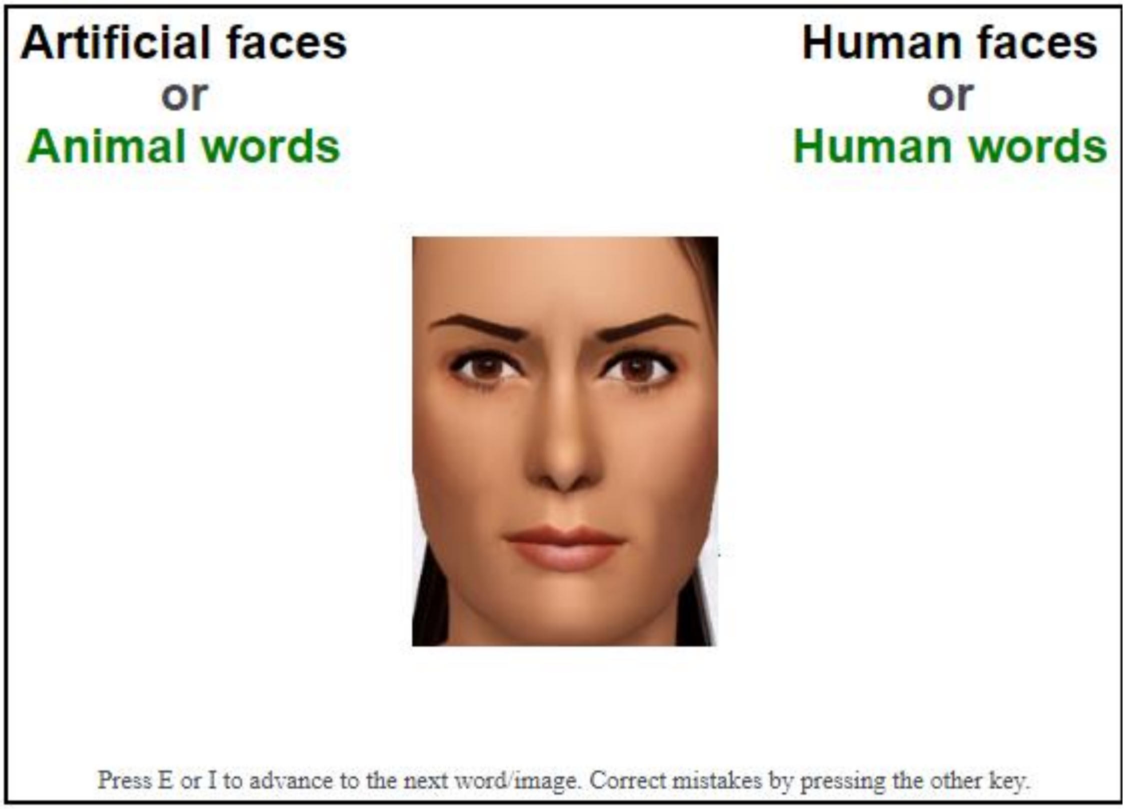


Figure 5. Reprinted from Wang et al. (in prep). Mean ratings of animacy at each level of exposure time (100ms, 500ms, and 1000ms) and for each face type (Human, android, and mechanical-looking). Error bars represent one standard error from the mean ratings.
*p < .01. **p < .001. ***p < .0001.

Appendix A: A Trial from a Congruent Block in Study 1



Press E or I to advance to the next word/image. Correct mistakes by pressing the other key.

Appendix B: Stimulus Words for the IAT in Study 1

Human words	**Animal words**
Wife	Pet
Maiden	Mongrel
Woman	Pedigree
Person	Breed
Husband	Wildlife
Humanity	Critter
People	Cub
Civilian	Creature
Man	Feral
Citizen	Wild

Appendix C: Stimulus Words for the IAT in Study 2

Human words	**Animal words**	**Automata words**
Wife	Alligator	Android
Maiden	Animals	Artificial
Woman	Beast	Automaton
Person	Cattle	Computer
Husband	Chimpanzee	Laptop
Humanity	Elephant	Machine
People	Kangaroo	Mechanical
Civilian	Mammals	Robot
Man	Platypus	Software
Citizen	Primates	Synthetic

Appendix D: Vignettes in Study 4

In Study 4, participants read the following three vignettes in which the target had a car accident and either fully recovered, lost experience, or lost agency. Participants then rated the level of mind the protagonist possesses.

Complete Human condition

Emma had a car accident and suffered from major injuries including damage to her brain. She was in a coma for a short time but woke up. Now, Emma is fully recovered. Her brain is fully functioning, and she has all of the mental capacities of a normal person.

Loss of Agency condition

Emma had a car accident and suffered from major injuries including damage to her brain. Although Emma did not die, she permanently lost her mental capacities to plan or make goals, to do things a normal person can do. However, she was still able to feel pain, pleasure or otherwise experience what a normal person can experience.

Loss of Experience condition

Emma had a car accident and suffered from major injuries including damage to her brain. Although Emma did not die, she permanently lost her mental capacities to feel pain, pleasure or otherwise experience what a normal person can experience. However, she was still able to plan or make goals, to do things a normal person can do.