Large language models have divergent effects on self-perceptions of mind and the attributes considered uniquely human

Jacobs, Oliver L.; Pazhoohi, Farid; Kingstone, Alan

Published in:
Consciousness and Cognition

DOI:
10.1016/j.concog.2024.103733

Publication date:
2024

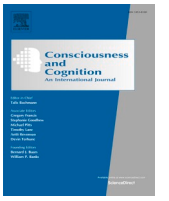
Document version:
Publisher's PDF, also known as Version of record

Consciousness and Cognition 124 (2024) 103733

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/yccog



Full Length Article

Large language models have divergent effects on self-perceptions of mind and the attributes considered uniquely human

Oliver L. Jacobs [a,*], Farid Pazhoohi [b], Alan Kingstone [a]

[a] Department of Psychology, University of British Columbia, Canada
[b] School of Psychology, University of Plymouth, United Kingdom

A B S T R A C T

The rise of powerful Large Language Models (LLMs) provides a compelling opportunity to investigate the consequences of anthropomorphism, particularly regarding how their exposure may influence the way individuals view themselves (self-perception) and other people (other-perception). Using a mind perception framework, we examined attributions of agency (the ability to do) and experience (the ability to feel). Participants evaluated their agentic and experiential capabilities and the extent to which these features are uniquely human before and after exposure to LLM responses. Post-exposure, participants increased evaluations of their agentic and experiential qualities while decreasing their perception that agency and experience are considered to be uniquely human. These results indicate that anthropomorphizing LLMs impacts attributions of mind for humans in fundamentally divergent ways: enhancing the perception of one's own mind while reducing its uniqueness for others. These results open up a range of future questions regarding how anthropomorphism can affect mind perception toward humans.

1. Introduction

The meteoric rise of highly capable AI chatbots such as ChatGPT and other large language model (LLM) applications has spurred calls to use psychological tools to investigate how interactions with these systems influence a number of perceptions (Shiffrin & Mitchell, 2023; Kosinski, 2023; Binz & Schulz, 2023). A variety of tools in psychology, cognitive science, and other related fields have emerged to study perceptions in the context of diverse human–computer interactions, which can be applied to LLMs. While this interest harkens back to an earlier period with chatbots such as ELIZA (Weizenbaum, 1966; Shum et al., 2018), the immense technological developments and widespread popularity of LLMs have reinvigorated the need to understand attitudes toward AI systems.

The novel capabilities of AI offer a particular opportunity to investigate the tendencies of people to anthropomorphize and its subsequent consequences. Anthropomorphism involves recognizing a spectrum of human-like traits, from seeing human-like shapes in the environment (commonly referred to as animacy; Bartneck et al., 2009), to attributing human-like qualities, both non-mental and mental, to nonhuman entities (Epley et al., 2007; Epley et al., 2008; Waytz et al., 2010).

The attribution of mental states has been distilled into two principal factors by Gray, Gray, and Wegner (2007): agency – the ability to do, think, and act morally; and experience—the ability to feel emotions, drives, and pleasure or pain. Their mind perception framework has been applied extensively to a wide variety of beings including humans, animals, robots, and AI systems (e.g., Wiese et al., 2017; Jacobs et al., 2022; Shank et al., 2019). Recently, this mind perception framework has also been applied to ChatGPT, revealing that even the briefest of exposures to ChatGPT can increase agentic (the ability to do) and experiential (the ability to feel) attributions to ChatGPT (Jacobs et al., 2023a). In other words, exposure can increase anthropomorphic attitudes toward LLMs—including those relating to the perception of mind.

* Corresponding author.
E-mail address: ojacobs@psych.ubc.ca (O.L. Jacobs).

https://doi.org/10.1016/j.concog.2024.103733
Received 26 February 2024; Received in revised form 16 July 2024; Accepted 25 July 2024
Available online 7 August 2024

In the initial work by Gray et al. (2007) one of the targets of mind perception was the self, such that individuals were asked to rate themselves on a variety of capacities of mind. People predictably rated themselves higher than nonhumans such as a dog, frog, or robot on both agency and experience. However, in their study, as in the mind perception literature in general, the measures of mind have tended to focus predominately on external mental attributions toward *others* rather than internal attributions toward the *self* (i.e., self-perception).

Though most investigations in the field have concentrated on one's perceptions of other minds (c.f. Jacobs et al., 2023), it is essential to acknowledge that our understanding of our own mind does not develop in isolation—it emerges within a comparative context. We define and perceive ourselves often in comparison to others (e.g., social comparison theory; Festinger, 1957; Gerber et al., 2018). The degree to which we do this with nonhumans is far less studied or conclusive, with past investigations often alluding to, but not directly addressing, self-comparisons with nonhumans (Kiesler et al., 2008; Waytz et al., 2010a; Brette, 2022). However, if we take as a working hypothesis that the external attributions of mind to nonhuman entities (i.e., anthropomorphism) can in turn affect our self-attributions —where humans not only attribute human-like qualities to nonhuman entities but also re-evaluate their own characteristics as a result—then there is a large gap in our understanding of anthropomorphic tendencies and self-perception. As interactions with increasingly sophisticated AI systems like ChatGPT become more prevalent, investigating this often-overlooked consequence of anthropomorphism becomes all the more important. Moreover, this interplay between self-perception and other-perception raises intriguing questions. For instance, how might our growing familiarity with LLMs impact our self-perception of mind (i.e. how we view our own minds) and what makes us human?

Approaches to this question have complex philosophical and religious roots that begin with Aristotle's numerous writings on the essence of what it is to be human with his emphasis on rationality (Barker & Stalley, 1995). Unsurprisingly, since that time a range of different viewpoints have emerged. For example, Charles Darwin famously noted that the difference in mind between man and higher animals, although great, is more a matter of degree than kind (Darwin, 1859). In contrast, other scientists such as Michael Gazzainga (2008) have leaned towards the metaphor of a phase shift in evolution, with the result being that it is nearly impossible to think of the minds of humans and animals as having the same constituent parts (Cacioppo & Patrick, 2008). Psychology and neuroscience also have a long history of searching for what makes the human brain unique, and by extension the human mind. A number of hypotheses have been put forth ranging from the high ratio of brain size to body size (Cairó, 2011), to the unusually large development of the cerebral cortex (Molnár et al., 2019), to the plasticity of human brains compared to our closest ancestors (Gómez-Robles et al., 2015).

Critically, the conceptualization of what makes the human mind unique has always been thought of in a comparative context first with animals and now, perhaps increasingly, with machines. Furthermore, a common thread between the disparate ideas that exist is that it is the cognitive abilities of the human mind that separate it from all else. It is our ability to think rationally, perform mathematical calculations, and create works of art that make us uniquely human. However, as AI encroaches on each of these capabilities, it is understandable that people begin to question if these agentic abilities 'to do' are what establishes the human mind as unique. Rather, people may begin to emphasize our experiential capabilities 'to feel' as crucial to what makes them in particular, and humans in general, unique. In short, it is not our ability to do that makes us special, but the way we feel and experience our world that makes us human. A key distinction is that when attempting to address these questions about self-perception and the degree to which certain qualities of mind are uniquely human, is that there are two approaches one can take. There is the loftier philosophical approach, and there is the empirical approach that seeks to survey folk psychology. The benefit of the mind perception framework is that it can be used for the latter to probe people's attitudes on these important questions (Jacobs et al., 2022).

The present study examines if exposure to LLM responses can influence self-perception of mind and the attributes of mind considered to be uniquely human. We predict that people will place greater emphasis on the experiential components of mind because as AI systems continue to become more sophisticated in their range of capabilities to do various tasks, the experiential capabilities of humans become comparatively larger than any agentic differences. To test this hypothesis, participants were recruited to fill out mind perception scales *rating themselves* and indicating the degree to which agency and experience are uniquely human, before and after brief exposure to ChatGPT prompts.

2. Methods

2.1. Participants

G*Power was used to conduct a power analysis to detect differences among groups for small effects ($d = 0.3$) with 95 % power. This suggested a sample size of at least 134 participants. To protect from participant dropout and exclusion, 150 participants took part via Amazon's Mechanical Turk (MTurk). 17 participants were excluded for rapidly clicking through the survey leaving 133 for the analysis. Of these participants, 75 were men, 57 were women, and 1 individual preferred not to identify. The mean age was 40.7 (SD=12.4) and all participants took part from IP addresses listed in the United States. This study was approved by the Behavioural Research Ethics Board of the University of British Columbia (H22-00572).

2.2. Procedure

After agreeing to participate following a brief description of the study that did not mention ChatGPT or LLMs, participants were asked to answer demographic questions related to their age and gender. Next, participants were asked to rate themselves on a modified 5-item Mind Survey (Gray et al., 2007; Tharp et al., 2017; Jacobs et al., 2023) which had a 5-point numerical scale from 'Not at all' (1) to 'A great deal' (5). These items were: 'How capable of exercising self-control are you?', 'How capable of remembering are you?', 'How

capable of feeling fear are you?', 'How capable of feeling self-pleasure are you?', and 'How capable of feeling hunger are you?'. A morality question was unintentionally omitted from the survey. The first 2 questions measured agency with acceptable reliability (α = 0.74), and the latter 3 questions measured experience with acceptable reliability (α = 0.71). Participants were then asked to rate themselves on agency (ability to do things) and experience (ability to feel things). These items were measured using a slider scale, with 0 indicating 'Not at all', and 100 indicating 'A great deal'. These questions were included in addition to the Mind survey in order to determine if the Mind survey questions would differ from simply asking for agency and experience ratings as is occasionally done (Will et al., 2021; Jacobs et al., 2022). Next, participants were asked to answer the question: 'To what degree do you think that agency (the ability to do things) makes humans unique compared to AI?' and "To what degree do you think that experience (the ability to feel things) makes humans unique compared to AI?" using sliding scales, from 0 to 100. After answering the previous questions, participants were presented with three real prompts provided by ChatGPT (GPT-4) along with its responses. The prompts were related to (a) the nature-nurture debate, (b) a tricky question about the number of feet that fit in a shoe, and (c) generating a cover letter for a janitorial position on the moon. These prompts were chosen to highlight the remarkable diversity of responses ChatGPT is capable of providing, from being educational to demonstrating reasoning and creativity. After reading these prompts, participants were once again asked to complete the above-mentioned 9 mind perception questions. The study took about 15 min for an individual to complete.

2.3. Data analysis

Data analyses and visualizations were conducted in R (v4.2.1; R Core Team 2021) using the packages tidyverse (Wickham et al., 2019) and papaja (Aust & Barth, 2018). Data are publicly available on OSF at https://doi.org/10.17605/OSF.IO/TYE3K.

3. Results

3.1. Mind survey scale

Items from the Mind survey (Gray et al., 2007) were averaged across the items corresponding to agency and the items corresponding to experience. These scores were then analyzed with a 2x2 within-subjects ANOVA with timepoint (pre versus post) and attribute (agency versus experience) as factors (see Fig. 1). There was a significant main effect of timepoint, $F(1,132) = 7.84$, $p = 0.006$, $\hat{\eta}^2_G = 0.002$, no main effect of attribute, $F(1,132) = 0.95$, $p = 0.332$, $\hat{\eta}^2_G = 0.002$, and the timepoint by attribute interaction was not significant, $F(1,132) = 1.00$, $p = 0.320$, $\hat{\eta}^2_G < 0.001$. This analysis indicates that agency scores and experience scores were significantly higher to a similar degree post-exposure ($M_{Agency} = 4.27$, $SD_{Agency} = 0.75$, $M_{Experience} = 4.35$, $SD_{Experience} = 0.71$) compared to pre-exposure ($M_{Agency} = 4.22$, $SD_{Agency} = 0.73$, $M_{Experience} = 4.26$, $SD_{Experience} = 0.74$).

3.2. Self-Ratings single item measures

Using the single-item mind perception assessments, a 2x2 within-subjects ANOVA was also conducted for the self-perception ratings made before and after exposure to the ChatGPT prompts (see Fig. 2). The main effects of attribute, $F(1,132) = 14.37$, $p < 0.001$, $\hat{\eta}^2_G = 0.015$, and timepoint, $F(1,132) = 4.64$, $p = 0.033$, $\hat{\eta}^2_G = 0.001$, were significant, and there was no interaction, $F(1,132) = 0.47$, $p = 0.500$, $\hat{\eta}^2_G < 0.001$. This analysis indicates that self-attributions of experience were significantly higher than that of agency ($M_{Experience} = 89.8$, $SD_{Experience} = 13.9$; $M_{Agency} = 86.3$, $SD_{Agency} = 16.4$), with the increase from pre-exposure ($M_{Experience} = 89.8$, $SD_{Experience} = 13.9$; $M_{Agency} = 86.3$, $SD_{Agency} = 16.4$) to post-exposure ($M_{Experience} = 91.1$, $SD_{Experience} = 13.4$; $M_{Agency} = 87.1$, $SD_{Agency} = 16.4$) being equivalent for both forms of attribution. Correlations between the Mind Survey scores and single-item responses for

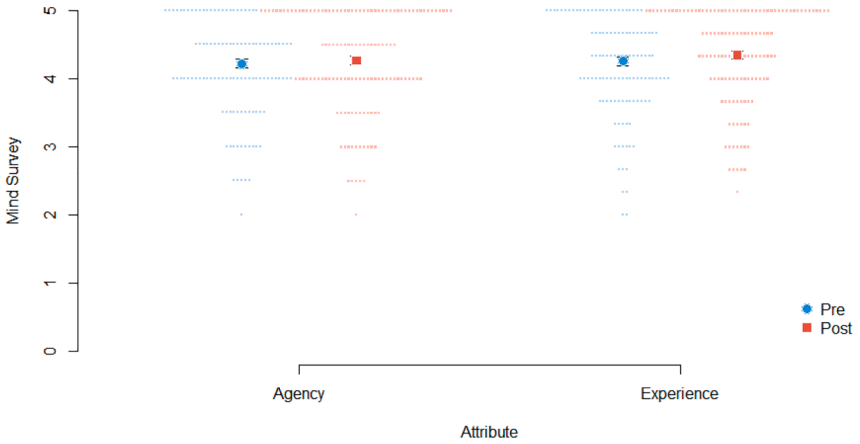


Fig. 1. Scaled self-ratings of mind perception pre-and post-exposure. Error bars indicate SEM.

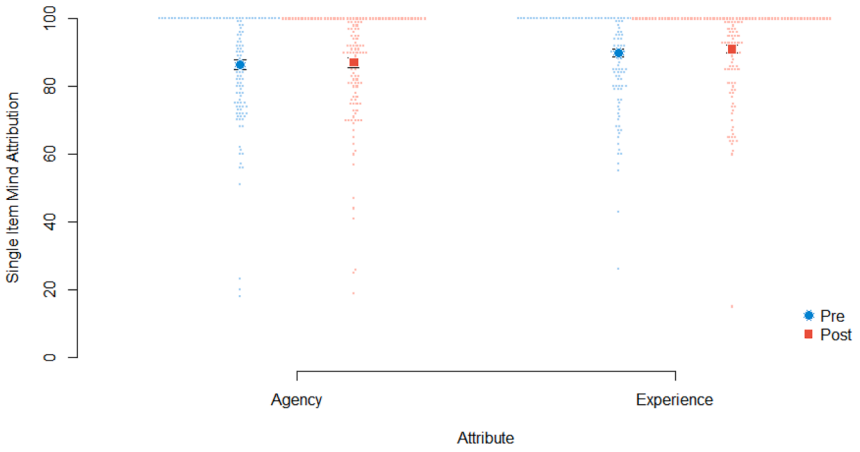


Fig. 2. Single-item self-ratings of mind perception pre-and post-exposure. Error bars indicate SEM.

agency ($r = 0.47$) and experience ($r = 0.57$) revealed they were strongly related.

3.3. Human uniqueness

A 2x2 within-subject ANOVA was also applied to the agency and experience (single items) attributions measuring the degree to which they are perceived to be unique to humans (see Fig. 3). As illustrated, experience ratings are higher overall than those for agency ($M_{Experience} = 86.9$ vs $M_{Agency} = 65.0$) with both falling a similar degree from pre-exposure ($M_{Experience} = 88.7$, $SD_{Experience} = 19.2$; $M_{Agency} = 67.1$, $SD_{Agency} = 28.7$) to post-exposure ($M_{Experience} = 85.1$, $SD_{Experience} = 22.7$; $M_{Agency} = 62.9$, $SD_{Agency} = 28.7$). This was confirmed statistically, with significant main effects for attribution, $F(1,132) = 65.61$, $p < 0.001$, $\widehat{\eta}^2_G = 0.161$, and timepoint, $F(1,132) = 7.62$, $p = 0.007$, $\widehat{\eta}^2_G = 0.006$, and no interaction, $F(1,132) = 0.03$, $p = 0.857$, $\widehat{\eta}^2_G < 0.001$.

These single-item measures indicate that self-perception attributions increase post-exposure, while attributions for human uniqueness decrease post-exposure. This divergence was confirmed statistically by entering question type (self-perception, human uniqueness) into the within-subject ANOVA, which returned a significant question type by timepoint interaction, $F(1,132) = 10.95$, $p = 0.001$, $\widehat{\eta}^2_G = 0.004$).

4. Discussion

Anthropomorphism involves the attribution of human-like qualities to nonhumans (Epley et al., 2007; Waytz et al., 2010a). Large language models (LLMs) with their increased capabilities and popularity offer a compelling opportunity to investigate the consequences of anthropomorphism on human perception. The mind perception framework is a popular psychological tool for measuring mental attribution that has previously been applied to examine self-perception and other-perception (Waytz et al., 2010b; Jacobs et al.,

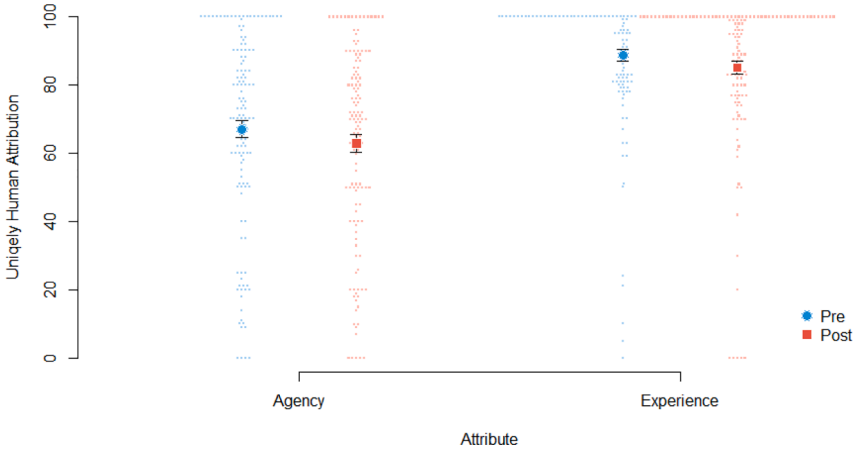


Fig. 3. Mind perception ratings (single item) pre-and post-exposure indicating the degree to which each attribute is uniquely human. Error bars indicate SEM.

2023b). This framework simplifies a wide variety of qualities of mind into their principal components: agency (the ability to do), and experience (the ability to feel) (Gray et al., 2007). Using this framework, we sought to probe self-perception and the features that individuals consider uniquely human before and after exposure to ChatGPT. Drawing from previous work that has used this mind perception framework (Jacobs et al., 2023a), we predicted that after exposure to ChatGPT, people will attribute greater experience to themselves and humans, in general, compared to agentic attributions.

Our investigation returned two major findings. First, self-perception was malleable as a function of exposure to ChatGPT prompts. Participants increased the amount of agency and experience they attributed to themselves after exposure to ChatGPT. We had predicted this effect to be especially salient for experience; that a comparable result occurred for agency was unexpected. These results could be a reflection of a shift in conceptual categories, which can be understood in terms of prototype theory (Rosch, 1973). Participants may consider themselves a better conceptual member (or prototype) of an entity—both in what they can do and feel—post-exposure to a LLM like ChatGPT. This is in effect also similar to social comparison theory (Festinger, 1954; Suls & Wheeler, 2013). However, unlike social comparison theory which focuses on how people compare themselves to other people, in the present case, the comparison concerns LLMs. This sort of downward comparison could explain the elevated self-perception of one's capabilities. Notably, these were consistent across both methods of measurement: single-item agency and experience assessments and the Mind survey.

Our second major finding was that, unlike self-perception, exposure to the LLM prompts *decreased* perceptions that agency and experience are uniquely human features. This divergence in results demonstrates that separating self-perception and other-perception (e.g. attitudes toward humanity) can reveal distinct effects. Moreover, it suggests that as programs like ChatGPT continue to develop and demonstrate more sophisticated abilities, individuals in turn may increasingly come to view agency and experience as less defining characteristics of humanness. Essentially, the implication is that there may be a change in social perspectives concerning the criteria of what defines uniquely human features. This fits neatly with the idea that if one's ability to think rationally, perform mathematical calculations, and create works of art forms a part of what makes us uniquely human, then as AI encroaches on each of these capabilities, then it follows that people will begin to question if traditional features of mind are unique. However, there was no interaction between facets of mind perception and exposure. Contrary to our hypothesis, people did not disproportionately think of agency as less unique than experience post-exposure. This is surprising considering that the developments in LLMs seem more related to increased agentic capacities compared to experiential capacities.

Importantly, although there was no interaction between exposure and facets of mind perception, people still viewed agency—the ability to do—as less unique to humans in comparison to attributions of experience. For example, many more people indicated experience as unique to humans at ceiling levels compared to agency (see Fig. 3). This fits with past work demonstrating high levels of attributing experience to other people and the general reluctance to ascribe experience to nonhumans (Gray et al., 2007). That said, people still considered agency as unique to humans and the difference with regard to experience was only relative, not in kind.

A number of limitations of the present study are worth considering. All participants took part remotely, and the degree to which the chosen prompts surprised participants was not measured. In light of our present results, measuring expectancy violation could shed light on the mechanisms driving the pre- and post-exposure effects. Furthermore, it is possible that the experience of the ChatGPT prompts may differ from interacting with ChatGPT in real time. The importance of real-time interactivity with LLMs remains an outstanding future research question, especially concerning mental attribution and anthropomorphism. Recent work has found that psychological distance, or the subjective experience of something being close or far from the self (Liberman et al., 2007), can be a mediating factor through which anthropomorphism positively affects evaluations of AI systems (Li & Sung, 2021) In the present study, the magnitude of effects was also relatively small. This may in part be because we adopted a conservative method for investigating differences, with the exposure manipulation to ChatGPT being relatively brief. It is possible that the short duration may have increased consistency bias effects among participants as they likely remembered their prior ratings and would desire to remain consistent. Understanding the impacts of long-term exposure as LLMs become more commonplace offers an additional fruitful direction for future research. Future directions may also include examining individual differences as they relate to anthropomorphism and self- and other-perception, given the wide inter-individual variation in mind perception noted elsewhere (e.g. Tharp et al., 2017; Gray et al., 2011).

Collectively the present findings add to the growing literature on the effects of human–computer interaction with LLMs. As the frequency and nature of LLM interactions increase, and with these interactions people become increasingly aware of its capabilities, it is crucial to consider how these interactions might shape attitudes toward ourselves in addition to wider societal attitudes (Turkle, 2017). The overall findings support our working hypothesis that anthropomorphism can in turn affect self-perception and attitudes toward other people's minds more broadly.

CRediT authorship contribution statement

Oliver L. Jacobs: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Farid Pazhoohi:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Alan Kingstone:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Investigation, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Support for this work was provided by a Discovery Grant to AK from the Natural Sciences and

Engineering Research Council of Canada (RGPIN-2022-03079).

Data availability

The data are available for public access on OSF: DOI 10.17605/OSF.IO/TYE3K

References

Aust, F., & Barth, M. (2018). Papaja: Create APA manuscripts with R Markdown.

Barker, E., & Stalley, R. F. (1995). *Aristotle: Politics*. Oxford University Press.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot., 1*, 71–81.

Binz, M., & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proc. Natl. Acad. Sci., 120*(6), Article e2218523120.

Brette, R. (2022). Brains as computers: Metaphor, analogy, theory or fact? *Front. Ecol. Evol., 10*, Article 878729.

Cacioppo, J., & Patrick, W. (2008). Are humans unique? *Nat. Neurosci., 11*, 1119.

Cairó, O. (2011). External measures of cognition. *Front. Hum. Neurosci., 5*, 108.

Darwin, C. (1859). *The origin of species*, (1872). London: reprinted by D. Appleton and Company.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychol. Rev., 114*(4), 864–886.

Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Soc. Cogn., 26*(2), 143–155.

Festinger, L. (1954). A theory of social comparison processes. *Hum. Relat., 7*, 117–140.

Gazzaniga, M. S. (2008). Human: The science behind what makes us unique. *Ecco/HarperCollins.*

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619.

Gray, K., Jenkins, A. C., Heberlein, A. S., & Wegner, D. M. (2011). Distortions of mind perception in psychopathology. *Proc. Natl. Acad. Sci., 108*(2), 477–479.

Gerber, J. P., Wheeler, L., & Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. *Psychol. Bull., 144*(2), 177.

Gómez-Robles, A., Hopkins, W. D., Schapiro, S. J., & Sherwood, C. C. (2015). Relaxed genetic control of cortical organization in human brains compared with chimpanzees. *Proc. Natl. Acad. Sci., 112*(48), 14799–14804.

Jacobs, O. L., Gazzaz, K., & Kingstone, A. (2022). Mind the robot! Variation in attributions of mind to a wide set of real and fictional robots. *Int. J. Soc. Robot., 14*(2), 529–537.

Jacobs, O., Pazhoohi, F., & Kingstone, A. (2023a). Brief exposure increases mind perception to ChatGPT and is moderated by the individual propensity to anthropomorphize. *PsyArXiv. March, 26*.

Jacobs, O. L., Pazhoohi, F., & Kingstone, A. (2023). Self-discrepancies in mind perception for actual, ideal, and ought selves and partners. *PLoS One, 18*(12), Article e0295515.

Kiesler, S., Powers, A., Fussell, S. R., & Torrey, C. (2008). Anthropomorphic interactions with a robot and robot–like agent. *Soc. Cogn., 26*(2), 169–181.

Kosinski, M. (2023). Theory of mind may have spontaneously emerged in large language models. *arXiv preprint* arXiv:2302.02083.

Li, X., & Sung, Y. (2021). Anthropomorphism brings us closer: The mediating role of psychological distance in User–AI assistant interactions. *Comput. Hum. Behav., 118*, Article 106680.

Liberman, N., Trope, Y., & Stephan, E. (2007). Psychological distance. *Social psychology: Handbook of basic principles, 2*(2), 353–383.

Molnár, Z., Clowry, G. J., Šestan, N., Alzu'bi, A., Bakken, T., Hevner, R. F., & Kriegstein, A. (2019). New insights into the development of the human cerebral cortex. *J. Anat., 235*(3), 432–451.

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rosch, E. H. (1973). Natural categories. *Cogn. Psychol., 4*(3), 328–350.

Shank, D. B., Graves, C., Gott, A., Gamez, P., & Rodriguez, S. (2019). Feeling our way to machine minds: People's emotions when perceiving mind in artificial intelligence. *Comput. Hum. Behav., 98*, 256–266.

Shiffrin, R., & Mitchell, M. (2023). Probing the psychology of AI models. *Proc. Natl. Acad. Sci., 120*(10), Article e2300963120.

Shum, H. Y., He, X. D., & Li, D. (2018). From Eliza to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering, 19*, 10–26.

Suls, J., & Wheeler, L. (2013). Social Comparison. In *Handbook of approach and avoidance motivation* (pp. 585–600). Psychology Press.

Tharp, M., Holtzman, N. S., & Eadeh, F. R. (2017). Mind perception and individual differences: A replication and extension. *Basic Appl. Soc. Psychol., 39*(1), 68–73.

Turkle, S. (2017). *Alone together: Why we expect more from technology and less from each other*. Hachette UK.

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspect. Psychol. Sci., 5*(3), 219–232.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends Cogn. Sci., 14*(8), 383–388.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM, 9*(1), 36–45.

Wickham, et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686.

Wiese, E., Metta, G., & Wykowska, A. (2017). Robots as intentional agents: Using neuroscientific methods to make robots appear more social. *Front. Psychol., 8*, 1663.

P. Will E. Merritt R. Jenkins A. Kingstone The Medusa effect reveals levels of mind perception in pictures Proc. Natl. Acad. Sci. 118 32 2021 e2106640118.