Realist, Anti-realist, or Agnostic? Exploring Folk-Ontological Stances toward Robot Minds

Sam Thellman
Dept. of Computer & Information Science, Linköping University
Linköping, Sweden
sam.thellman@liu.se

Klara Bergsten
Dept. of Computer & Information Science, Linköping University
Linköping, Sweden
klabe908@student.liu.se

Edoardo Datteri
Dept. of Human Sciences for Education, University of Milano-Bicocca
Milan, Italy
edoardo.datteri@unimib.it

Tom Ziemke
Dept. of Computer & Information Science, Linköping University
Linköping, Sweden
tom.ziemke@liu.se

Abstract

People routinely attribute mental states such as beliefs, desires, and intentions to explain and predict others' behavior. Prior work shows that such attributions extend to robots, yet it remains unclear what people assume about the *reality* of the states they attribute to them. Building on recent conceptual work on folk-ontological stances, we report a pilot study measuring realist, anti-realist, and agnostic stances toward robot minds. Using a questionnaire (N = 66), we assessed stances toward today's robots and robots in principle, and examined stance rigidity through a reflection-and-reassessment design. Results show stronger anti-realist tendencies for today's robots than for robots in principle. Stances were largely rigid across reflection. Notably, participants did not hold a uniformly non-realist view but expressed a diversity of folk-ontological stances, including substantial proportions of agnostic and realist responses. This heterogeneity highlights the need for measurement tools that move beyond binary measures and capture nuance in folk-ontological reasoning. Future work will expand stance options to include finer-grained realist and anti-realist variants and recruit cross-cultural samples to assess variation across populations.

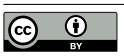
CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; • **Applied computing** → *Psychology*.

Keywords

folk ontology, mental state attribution, belief vs. acceptance

ACM Reference Format:
Sam Thellman, Klara Bergsten, Edoardo Datteri, and Tom Ziemke. 2026. Realist, Anti-realist, or Agnostic? Exploring Folk-Ontological Stances toward Robot Minds. In *Companion Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction (HRI Companion '26), March 16–19, 2026, Edinburgh, Scotland, UK*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3776734.3794410



HRI Companion '26, Edinburgh, Scotland, UK

ACM ISBN 979-8-4007-2321-6/2026/03
https://doi.org/10.1145/3776734.3794410

1 Introduction

People tend to interpret observed behavior as driven by underlying mental states such as beliefs, desires, and intentions. This practice is not limited to humans: people also attribute mental states to animals, fictional characters, and artificial intelligent systems, including robots [4, 5, 7]. Mental-state attributions are often accompanied by underlying beliefs about whether the attributed states correspond to something *real* within the agent to whom they are attributed, or whether they instead function merely as convenient ways of conceptualizing that agent's behavior. Such assumptions are commonly referred to as *folk-ontological* stances [3, 8].

In the case of humans or non-human animals, people typically presume that when we attribute mental states to the agent, those states track genuine internal states (for example, brain states in the organism) — thereby adopting a *realist* stance. In other cases, such as when we attribute mental states to collective entities or to fictional characters (e.g., "The Roman Empire feared retaliation from the Germanic tribes after the disaster in the Teutoburg Forest" or "Donald Duck thinks the Chipmunks want to steal his pancakes"), the assumption often is that the attributed mental states do *not* correspond to anything real in the agent in question — reflecting an *anti-realist* stance (called eliminativist in [3]). Finally, there can be cases where people lack a settled assumption about whether the mental states are real or not — in which case we may say they adopt an *agnostic* stance.

As noted by Datteri [3], little is known about people's folk-ontological stances toward robots and how these stances influence interaction. Prior work has found that people tend to deny robot minds when asked explicitly (e.g., "Do you think [the robot] has a mind?") [1]. However, commonly used yes/no measures[1] are restrictive, leaving no room for agnostic positions, and often ambiguous about whether they concern today's robots or robots in principle. A further issue is that responses may reflect either *beliefs* or more context-sensitive *acceptances*, which can have different downstream effects on attitudes and behavior [2]. To address these

[1] Yes/no measures also have a distinct advantage over graded measures such as Likert scales: they avoid implicitly treating "having a mind" as a quantity that an agent can possess to a greater or lesser extent—a dubious and theoretically contentious assumption [6].

limitations, the current project aims to systematically investigate folk-ontological stances toward robots empirically.

In this pilot study, we operationalized folk-ontological stances by presenting participants with short descriptions of realism, anti-realism, and agnosticism, and asked them to indicate alignment. We focused on three research questions:

RQ1 What are people's folk-ontological stances (realist, anti-realist, or agnostic) toward *today's robots* having a mind?
RQ2 What are people's folk-ontological stances (realist, anti-realist, or agnostic) toward the possibility of *future robots* having a mind?
RQ3 How rigid or flexible are people's folk-ontological stances toward robots?

2 Background

2.1 Mental State Attribution to Robots

Humans readily interpret robot behavior using mentalistic concepts, especially when robots display autonomy, goal-directedness, or social cues [7]. Nevertheless, as Datteri [3] points out, it is not obvious that attributing a mental state to another agent can be equated with the linguistic act of uttering a mentalistic sentence about that agent. Datteri defines "attribution" in terms of people's beliefs about robots rather than their utterances. Therefore, attributing a mental state S to a robot means believing that the robot possesses that state.

Building on this definition of the concept of "attribution", Datteri [3] defines several folk-ontological stances toward robots' minds. These are conceived as ordinary people's beliefs about whether attributed mental states are real or, for instance, merely useful fictions. Although a rich taxonomy of folk-ontological stances can be identified, our pilot study adopts the coarse tripartite distinction between realism, anti-realism and agnosticism to ensure empirical tractability. A realist folk-ontological stance is adopted when one believes that a robot genuinely possesses mental states. An anti-realist stance consists of believing that robots do not have mental states (this stance is referred to as "eliminativism" in [3]). Agnosticism amounts to an absence of belief or a state of uncertainty about the reality of a robot's mind.

Datteri defines these folk-ontological stances solely in terms of beliefs. However, it is important to note that people's verbal responses about robot minds may reflect either beliefs (an involuntary mental state in which somebody "feels" a proposition to be true) or acceptances (a voluntary stance adopted for the purpose of reasoning or action, whereby a proposition is treated as true even if it is not fully believed). As pointed out in [2], beliefs and acceptances differ in many interesting ways, including their logical behavior. Since beliefs develop spontaneously in people, they are not usually revised based on new evidence or logical reasoning. Acceptances, on the other hand, are voluntarily and consciously held, and people tend to revise them based on new evidence to preserve consistency through logical inference. Consequently, beliefs and acceptances can lead to different behavioral and attitudinal consequences. For this reason, this study investigated people's folk-ontological stances both in terms of their beliefs and acceptances about robots' minds and incorporated a reassessment phase to probe response rigidity/flexibility as an initial indicator of belief-based versus acceptance-based stances.

3 Method

3.1 Design and Procedure

We collected questionnaire data in four parts:

(1) Background information (demographics and experience).
(2) Views on robots' thinking abilities, split into (2A) thinking in today's robots and (2B) thinking in robots in principle, including future robots.
(3) Reflective questions about the nature of thinking and mind.
(4) Opportunity to reassess initial views after reflection.

Participants were told that no expert knowledge was required, that there were no right or wrong answers, and that we were interested in their intuitive views. All provided informed consent.

Operationalization note. As a study delimitation, we operationalized ontological stance-taking toward robot minds using questions excluding the terms "mind" and "mental states", which ambiguously refer to aspects of the mind related to both phenomenological experience (e.g., having feelings and sensations) and intentional agency (e.g., having beliefs and intentions). Instead, we phrased stance questions in terms of *thoughts and intentions*[2], which avoid this some of this ambiguity, referring to intentional states/processes without necessarily invoking the concept of phenomenal experience.

3.2 Participants

Sixty-seven Swedish-literate participants (18–60 years, median age 27) were recruited via convenience sampling and completed the survey in Swedish[3]. One participant failed a control question and was excluded, leaving $N = 66$. Gender distribution was balanced (34 women, 33 men). Occupation was similarly balanced (35 employed, 32 students). Educational levels were varied: 22.4% held a Master's degree or higher, 32.8% a Bachelor's degree, 10.4% post-secondary vocational/folk high school, 32.8% high school, and 1.5% primary school.

Self-rated experience with robots and AI was: none for 4.5% of the participants, limited 61.2%, intermediate 26.9%, and high 7.5%.

3.3 Measures

Part 2A: stance toward today's robots. Participants rated agreement (5-point Likert: Strongly Disagree–Strongly Agree) with three position statements and selected the best match:

- **Realism**: "Today's robots have thoughts and intentions."
- **Anti-realism**: "Today's robots lack genuine thoughts and intentions."
- **Agnosticism**: "I don't know whether today's robots have thoughts and intentions."

Confidence was rated (1–5), with optional free-text justification.

Part 2B: stance toward robots in principle/future robots. The same structure was used, reframed to concern robots *in principle*, including future robots:

[2]Translated from Swedish "tankar och avsikter".
[3]The Swedish survey questions and items were freely translated into English for this report.

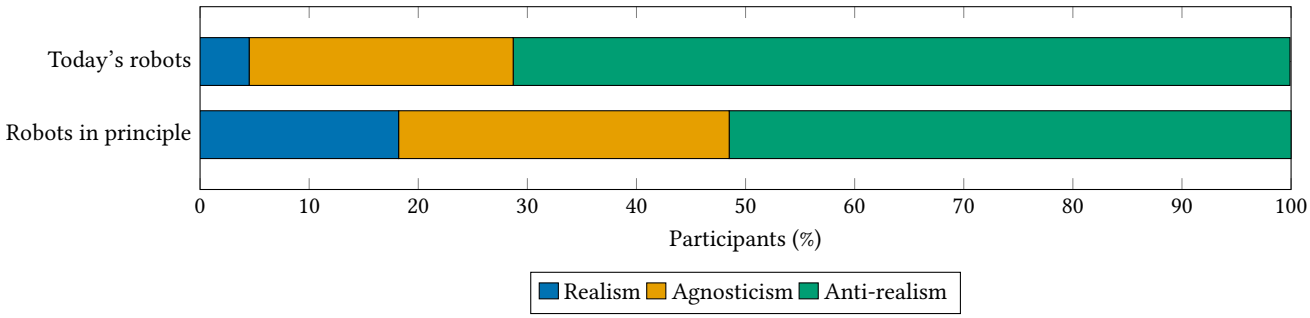


Figure 1: Proportion of participants aligning with each stance for today's robots vs. robots in principle.

- **Realism**: "Robots can have thoughts and intentions".
- **Anti-realism**: "Robots cannot have thoughts and intentions".
- **Agnosticism**: "I don't know whether robots can have thoughts and intentions".

Part 3: reflection items. Participants rated agreement with 15 claims about thinking, mind, biology, consciousness, and machine computation (plus a control item). These items served only to induce reflection rather than constitute primary outcomes (see Appendix A).

Part 4: reassessment. Participants re-selected realism/anti-realism/agnosticism for robots in principle and re-rated confidence, with optional explanation.

4 Results

4.1 Demographic Effects

Chi-square tests indicated no significant association between self-rated robot/AI experience and stance choice at any measurement point (all $p > .05$). Likewise, no significant associations were found between education level and stance choice (all $p > .10$). Given space constraints, we do not detail these analyses further.

4.2 Stances Toward Today's Robots

A large majority (71.2%, $n = 47$) selected anti-realism for today's robots, 24.2% ($n = 16$) selected agnosticism, and 4.5% ($n = 3$) selected realism. Confidence was generally high across stances (medians near 4/5), with the greatest variance among agnostics. Due to the small realist subgroup, we did not run inferential tests on confidence differences.

Free-text explanations from anti-realists centered on robots as programmed or algorithmic systems lacking biological bases or consciousness. Several participants allowed that robots could *simulate* thinking, but claimed the underlying processes differ fundamentally from human thinking. Agnostics typically cited uncertainty about AI potential, limitations in knowledge, or ambiguity in defining "thinking." Realists emphasized the role of definitions: depending on how thinking is defined, robots could count as thinkers.

4.3 Stances Toward Robots in Principle

When considering whether robots can have thoughts and intentions in principle, anti-realism was less common (51.5%, $n = 34$) compared to the case of today's robots, agnosticism was more common (30.3%, $n = 20$), and realism was also more common (18.2%, $n = 12$). Confidence medians were 4 for anti-realism and agnosticism, but lower for realism (median 2.5). A Kruskal–Wallis test found no significant confidence differences among stances ($p = .149$).

Explanations for anti-realism echoed those for today's robots. Agnostics emphasized open-ended technological development and definitional uncertainty. Realists focused on the possibility that future science could replicate or reinstantiate biological mechanisms underlying thinking.

4.4 Comparison Between Today's Robots vs. Robots in Principle

Comparing stance selections across Parts 2A (today's robots) and 2B (robots in principle) revealed systematic differences (see Figure 1). Anti-realism was endorsed by fewer participants when evaluating robots in principle ($n = 34$) than when evaluating today's robots ($n = 47$), a statistically significant difference ($p = .002$). Realism, in contrast, was endorsed by more participants for robots in principle ($n = 12$) than for today's robots ($n = 3$), also a statistically significant difference ($p = .003$). Agnosticism was somewhat more common in judgments about robots in principle ($n = 20$) compared to today's robots ($n = 16$), although this difference was not statistically significant ($p = .301$). Overall, participants appeared less inclined to categorically deny the possibility of robot minds when considering robots in principle, suggesting a degree of openness to future robot mentality even among those initially aligning with anti-realism.

4.5 Rigidity vs. Flexibility After Reflection

We compared stance choices before and after the reflection component (Parts 2B and 4). Most participants maintained the same stance across the two assessments: of the 34 anti-realists prior to reflection, 31 remained anti-realist; of the 12 realists, 10 continued to identify as realist; and of the 20 agnostics, 19 retained their agnostic stance. Wilcoxon signed-rank tests showed no significant differences between pre- and post-reflection stance distributions for any category (all $p > .40$). Confidence levels displayed a similar pattern of stability, with only minor non-significant variations.

Taken together, these results suggest that folk-ontological stances in this pilot may align more closely with belief-like commitments, which are characteristically resistant to revision even when inconsistencies arise. Nevertheless, future work using multiple diagnostic markers of belief versus acceptance will be needed to examine this more rigorously.

5 Discussion

This pilot provides the first systematic empirical snapshot of folk-ontological stances toward robot minds. Three main findings stand out.

First, participants overwhelmingly endorsed non-realism for *today's* robots, consistent with prior explicit-denial findings, but now quantified alongside a substantial agnostic minority. The agnostic option appears important: a quarter of respondents preferred uncertainty over categorical denial or affirmation, supporting the claim that yes/no measures underestimate ambivalence.

Second, when asked about robots *in principle*, participants were significantly more open to the possibility of robot minds. Non-realism decreased, realism increased, and agnosticism remained prevalent. People thus seem to separate current technological limitations from future conceptual possibility, and our two-timepoint stance measure captures this distinction more cleanly than previous work.

Third, stances were largely rigid across the reflection-and-re-assessment manipulation. One interpretation is that folk-ontological stances function more like context-insensitive beliefs than context-sensitive acceptances. However, our reflection phase was short and generic; stronger manipulations or richer interactional contexts might reveal more flexibility.

5.1 Limitations and Future Work

As a pilot, this study has clear limitations. Our operationalization used "thoughts and intentions" in questions about robot minds, which may be interpreted differently across participants. Also, we used only coarse realist/anti-realist/agnostic categories, rather than the richer space of variants articulated by Datteri [3]. The sample was Swedish-speaking and convenience-based, limiting generalizability. Finally, belief vs. acceptance distinctions were only indirectly probed through stance rigidity.

In follow-up work, we will (1) operationalize folk-ontological stances toward minds without relying on "thoughts and intentions," (2) allow participants to align with finer-grained realist and non-realist variants, (3) recruit a global sample to examine cross-cultural differences, and (4) include additional direct markers of belief versus acceptance.

6 Conclusion

Measuring folk-ontological stances toward robot minds requires moving beyond binary questions. Our pilot suggests that most people are anti-realist about today's robots, but many are agnostic or realist about future robots, and that these stances are relatively stable under brief reflection. These findings motivate larger cross-cultural studies and improved operationalizations to understand how folk ontology shapes human–robot interaction.

A Reflection Items

Part 3 of the questionnaire invited participants to reflect on the nature of thinking and mind. Participants indicated their agreement with each statement on a 5-point Likert scale (Strongly Disagree – Disagree – Neutral – Agree – Strongly Agree). These items were included to induce reflection and were not analyzed as primary outcomes.

(1) Science has answered the question: "What is thinking?"
(2) To judge whether something thinks, we must know what thinking is.
(3) Science has answered the question: "Can robots think?"
(4) Thinking is a biological process.
(5) Unlike objects such as chairs or planets, thoughts occupy no physical space.
(6) *Control question:* Please select "Agree" for this statement.
(7) Animals can think.
(8) Insects can think.
(9) Bacteria can think.
(10) I know that I think.
(11) I know that other people think.
(12) Thinking requires a nervous system.
(13) Consciousness and thinking are the same thing.
(14) It is possible to think without being aware of thinking.
(15) There is a fundamental difference between human thinking and machine computation.
(16) Consciousness is an illusion.

References

[1] Jaime Banks. 2020. Theory of mind in social robots: replication of five established human tests. *International Journal of Social Robotics* 12, 2 (2020), 403–414.
[2] L Jonathan Cohen. 1989. Belief and acceptance. *Mind* 98, 391 (1989), 367–389.
[3] Edoardo Datteri. 2025. Folk-Ontological Stances Towards Robots and Psychological Human Likeness. *International Journal of Social Robotics* 17, 2 (2025), 257–276.
[4] Morton Ann Gernsbacher, H Hill Goldsmith, and Rachel RW Robertson. 1992. Do readers mentally represent characters' emotional states? *Cognition & Emotion* 6, 2 (1992), 89–111.
[5] Tyler J Kasperbauer. 2017. Mentalizing animals: implications for moral psychology and animal ethics. *Philosophical Studies* 174, 2 (2017), 465–484.
[6] John R Searle. 1992. *The rediscovery of the mind*. MIT press.
[7] Sam Thellman, Maartje De Graaf, and Tom Ziemke. 2022. Mental state attribution to robots: A systematic review of conceptions, methods, and findings. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 4 (2022), 1–51.
[8] Kara Weisman. 2022. Extraordinary entities: Insights into folk ontology from studies of lay people's beliefs about robots. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 44.

Received 2025-12-08; accepted 2026-01-12