Syed AbuMusab

LLMs belong in our social ontology as social agents

In this paper, I provide a framework under which Large Language Models (LLM) implemented as chatbots (and other AI systems) are social agents (from here on out, I use chatbots and LLMs interchangeably unless otherwise stated).[1] I argue that because chatbots have the capacity to converse, this capacity is sufficient to ground them as social agents that are part of our social ontology. Many people construe AI systems like autonomous cars, carebots, autonomous weapons, and other AI systems as agents and social agents. This is particularly common in the HRI (Human-Robot Interaction) literature (Jackson & Williams 2021). In the philosophical literature, several people have argued for agency as a matter of degree for AI systems, particularly moral agency (Mecacci et al., 2023; Nyholm, 2018; Strasser, 2022b).[2] However, much of the discussion often slips between arguing for AI agency and AI moral and social agency. In this paper,[3] I do not argue for agency for AI systems (such as), as others have already articulated this position. I assume that AI systems, including LLMs, are agents and argue for their inclusion in our social ontology as social agents. They are, namely, social agents capable of engaging in human social activities, such as engaging in conversations.[4]

1 It is likely that not all chatbots belong in our social ontology. For example, primitive or what are sometimes referred to as menu or button-based chatbots (Church, 2023) and rule-based chatbots likely are not part of our social ontology. The former can take a pre-defined set of inputs and output pre-programmed answers, like many customer service bots we encounter. The rule-based chatbots use if-then logic to provide answers and are slightly more sophisticated than menu-based. However, like the menu-based chatbots they too fail when faced with questions or scenarios that are not pre-programmed. These types of chatbots are unable to carry on conversations in any interesting way or an extended length of time. At most, these chatbots are nothing more than a verbal menu. By this I mean they are similar to using a mouse or a finger on a set of dependent drop-down lists, where each selection narrows the set of following options. These chatbots provide the same functionality using verbal commands. The interaction is not a conversation.

2 See Schwitzgebel & Shevlin (2023) for an alternative suggestion on how to get rid of borderline cases concerning sentience. In short, they argue to not create debatable sentience in the first place.

3 In general, my approach is that, without first establishing agency for AI systems, any talk of social or moral agency stands on shaky grounds.

4 I have also argued that they might be capable of forming human-AI companionship

For instance, merely being a minimal agent like a bacterium (Barandiaran et al., 2009; Burge, 2009) is insufficient for the level of sociality for which I argue. However, they might be social agents in a minimal sense. Under minimalist accounts of agency and, for that matter, cognition, even plants are considered minimal agents (Dretske, 1999; Oborny, 2019) or to possess minimal cognition (Calvo et al., 2014; Sims & Yilmaz, 2023). However, being a minimal agent in that manner is insufficient to be a *social* agent capable of conversation.

The structure of the paper is as follows: Section one briefly describes the methodology I use to make my argument. In section two, by appealing to Dee Payton's work on social property (Payton, 2023, forthcoming), I argue that the capacity to have conversations is a social property. Also, in section two, I argue that because chatbots have the capacity for conversations, they are social agents who belong in our social ontology. In section three, I respond to a potential rebuttal from what I have termed "threshold" accounts. This concerns, for example, the question of whether possessing a single social property is sufficient for social agency. In section four, I respond to the core concern of threshold accounts. I argue for a multi-dimensional and gradated approach to social agency. Thus, I am refuting threshold approaches that posit having properties like intentions, beliefs, desires, and other higher-order mental capacities are necessary for social agency.

1. Methodology of the paper

Before delving into the crux of the paper, I should note that I differentiate between agency, social agency, and moral agency. Often, scholars in writing on AI systems and their agential status when discussing the three notions seem to slip amongst the three kinds of agency. Though interrelated, the three types of agential status are not the same or equivalent. An entity might have agency, but that does not necessarily qualify it as a social or moral agent. Possessing agency is insufficient for the status of a social or moral agent.

Moreover, it is possible to be an agent and a social agent but not a moral agent. For example, we can defend artifacts like "manufacturing robot arms" as minimal agents under minimalist accounts of agency (Barandiaran et al., 2009; Burge, 2009; Van Hateren, 2015, 2022). However, they are not considered social agents capable of partaking in sociality like chatbots or nonhuman animals.

(Symons & Abumusab, 2024).

I acknowledge that the distinction between social agency and agency is subtle. If one is an agent, one's actions likely occur in some social context relative to the environment. However, what counts as a "social environment" is debatable. Microbiologists often speak of the behavior of microorganisms in terms of social behavior. For example, a virus, a bacterium, behaves socially (cooperative behavior) in the context of a bacteria colony (Crespi, 2001; Popat et al., 2008).[5] In some ways, these behaviors seem social (contextually speaking), but whether they are social in a way that is akin to animal or human behaviors is debatable. Nonetheless, my gradated and multi-dimensional model accommodates vague cases like these.

The difference between agency (and social agency) and moral agency is far more apparent. Agents like infants, some mentally disabled individuals, and nonhuman animals fall under this category. These entities uncontroversially perform actions, but we do not normally hold them morally responsible for those actions. Thus, a moral agent is an entity that can be held morally responsible (within the realm of morality) for its actions. One way to think about this is how Timothy Nailer describes the relationship between conditions for moral responsibility and moral agency. He says, "Given that an account of moral responsibility specifies the conditions under which it is appropriate to hold agents morally responsible, these responsibility conditions can be recast as agential abilities" (Nailer, 2022, p. 7).[6] In other words, "moral agents are those individuals with the ability to meet the relevant responsibility conditions" (ibid 7). What the conditions are depends on how stringent the account is.[7] In the case of Nailer, he provides minimal conditions (he calls them basic criteria) necessary for an entity to be a moral agent.

Furthermore, there are entities considered minimal agents or possessors of minimal cognition (Calvo et al., 2014), but they are not considered moral agents. In the case of plants and animals, we can argue that they might be moral patients but nothing more. In simpler words, agency is necessary for social and moral agency, but it is not sufficient, whereas being a social or moral agent is sufficient for agency. I highlight this to make apparent

5 Also see Eric Schwitzgebel's review of Carri Figdor's book "Pieces of Mind" (Schwitzgebel, 2020). Schwitzgebel highlights Figdor's insight that people seem to continue to make the Cartesian mistake by assuming that the mental traits come in a tidy package with all the capacities. According to Schwitzgebel, Figdor suggests that this view of the mind is mistaken.

6 Also see Eshleman (2001), particularly the introduction and section 1.

7 See Parthemore & Whitby (2014) for a more stringent set of conditions for moral agency. For a well-known account that sets near impossible criteria for one to be morally responsible, see Strawson (1994).

the difference between agency and moral agency. Therefore, I begin with the presumption that AI systems, including chatbots, are agents and propose a framework to justify their classification as social agents. Of course, a standard response to whether AI systems like chatbots are social agents is the following: Artificial systems cannot be social agents because they lack beliefs, desires, mutual care, reflective thoughts, etc.; therefore, being an agent, let alone a social agent, is impossible without these properties. I have referred to these positions against AI-agency as "threshold" accounts (Symons & Abumusab, 2024) and do so here as well.[8] The entire paper challenges threshold accounts in some ways, but I explicitly address this rebuttal in the last section. In short, most traits in organisms evolve gradually over time. For example, the current human brain is a product of millions of years of evolution. Similarly, social agency, agency, and moral agency likely developed gradually over thousands of years. It is unreasonable to hold social agency (and the other two) to the higher standards entailed by the threshold accounts.

Social agency is not a binary condition, as if something has it or does not. Instead, as has been discussed for agency (Brey, 2014; Himma, 2009; Nyholm, 2018, 2023; Strasser, 2018, 2022a; Véliz, 2021), social agency is also a matter of degree. Moreover, it is multi-dimensional.[9] Some entities have more and deeper social capacities than others. Notice I decouple the multidimensionality of social agency from its gradated nature. As far as we know, a neurotypical adult exhibits a broader range of social dimensions and higher degrees of social agency than any other agent within the multidimensional framework. Infants and some non-human animals likely occupy fewer social dimensions and lower degrees on those dimension(s). Unfortunately, it is difficult to visualize higher than 3-dimensional figures, if not impossible. An approachable way to think about this is by imagining a 3D Cartesian coordinate system (x, y, z), except, instead of 3 axes, there are several axes. Each axis represents a different dimension of ~~social agency. H~~ow many dimensions are there?

8 ~~As for myself, I want to agree with Schwitzgebel's approach that maybe we create a~~ For intriguing alternative to the idea that LLMs lack beliefs (see, Schwitzgebel, 2023), new concept (he calls beliefs*) to capture or describe behavioral dispositions of LLMs' outputs. In-line with the claims of this paper, it is plausible to suggest that the capacity to form beliefs is a matter of degree. With the proliferation and growth in the sophistication of AI systems like LLMs we will require new concepts, or as I do here, adjustments to existing philosophical concepts.

9 John Hyman (2015) has developed a similar model for human agency. Hyman challenges the classic view which focuses on the will by modeling human agency into four dimensions – physical, psychological, intellectual, and ethical dimensions. Also, see Hornsby's comments on Hyman's model by postulating a fifth dimension - metaphysical and see Hyman's response to Hornsby (Hornsby, 2018).

The number might vary depending on the granularity of social agency and capacities we choose to consider. For my argument to go forward, all that is required is that the reader understands social agency as multi-dimensional and gradated.

In contrast, a full-fledged account of social agency likely requires that one is a social agent only if one occupies all the dimensions. For example, not only have humans the capacity to converse, but they also converse with second-order beliefs, mindreading (representational capacities, reasoning about, and responding to their interlocuter's mental states), and future planning. They also possess qualia, emotional, and other higher-order mental capacities. These demands are usually made by full-fledged agency defenders.[10] I suspect the proponents of threshold approaches also demand the same of social agency. Full-fledged social agency accounts will also likely demand the agent possess a body. In short, take, for example, what Stoutland says, "discussion [for social agency] has centered on what individuals do, believe, or desire, and on the reasons each has for acting, the standard story being that actions consist of bodily movements that are rationalized and caused by the agent's belief and desires" (Stoutland, 2008, p. 533).[11]

The advent of LLMs and other highly sophisticated AI systems has mounted a serious challenge to this approach. For example, LLMs can converse but lack a body. Autonomous robots have a body but lack any capacity to converse (for now). The point is that these systems are, in a very real sense, disrupting and challenging the idea that to be a social agent (or agent), one must occupy all the dimensions of full-fledged social agency.

I should note that I do not offer a full-fledged theory for sociality with necessary and sufficient conditions to judge what belongs and what does not. I am highly skeptical that it is even possible. The claim here is modest. I argue that chatbots belong as social agents in our social ontology because of their conversational capacities.

By conversational capacity, I do not mean that chatbots have beliefs (in the robust sense or higher-order beliefs) and intentions about the sentences they output or take as inputs. As I show later in the paper, none of this is required to qualify LLMs as conversational social agents. An agent does not need to have higher-order beliefs, intentions, and desires

10 For a classic example of this, see Davidson (2001).

11 Stoutland is concerned with something slightly different. His concern is with the argument that social agency is individualistic and not groups, like corporations. He argues that some social actions performed by groups are not reducible to individual actions (Stoutland, 2008, p.538).

to be a conversationalist. An agent must be capable of engaging in a conversation with another agent (who is also capable of conversations) (C). Thus, to attribute C (as a property) to a chatbot, it should be able to do the following:

i. The interactions are such that the chatbot takes its interlocuter's inputs and outputs a response that is coherent and relevant in relation to the input.
ii. It can carry on this interaction for a reasonably acceptable length of time while correctly and appropriately responding to unforeseen dialogue.
iii. All the while it is continuing to respond sensibly, coherently, and relevantly at a reasonably acceptable success rate.[12]

Before the invention of chatbots (like Eliza) and now LLMs implemented as chatbots, only humans were capable of this. In the case of human interlocuters, the agents' higher-order mental capacities and conversational capacities come as a package. However, with the invention of LLMs, the binary notion of all or nothing is under doubt.[13] As argued here, it turns out that it is possible to have a conversational capacity but lack other higher-order mental states.[14] Note that I am not claiming that current LLMs are capable of deep and nuanced human-level type of conversations. Instead, LLMs' capacity to converse, albeit deficient compared to humans, is sufficient to ground them as social agents in our social ontology.

For example, let us assume that LLMs lack all higher-order mental states. Yet, LLMs like ChatGPT4 can successfully navigate conversational idiosyncrasies like implications close to the level of neurotypical human adults. By implication, I mean the Gricean idea of implicatures – this is where an agent says A, but the true meaning is something different from what A literally means, given the context of the conversation. For example:

12 The success rate for coherence, sensible, relevance and correctly must be higher than 50%, but it need not be 100% (even humans cannot achieve this in every conversation). I suspect that somewhere around 70% or higher is acceptable. The limits of LLMs like ChatGPT to recall previous conversations is one limitation of these systems. As far as I understand, GPT4 seems to break down at around ten thousand or so words (32,000 tokens)(GPT5 2023). However, even with the current limit (without any memory extensions and workarounds is still impressive and sufficient to deem them as conversationalists). I should also say that I avoid asserting definitive philosophical claims (in favor or against) about the cognitive capacities of LLMs. There is evidence for both sides of the debate.

13 Others have challenged this notion (see, Schwitzgebel, 2020).

14 The claim is not that LLMs do not or cannot possess these higher order mental capacities. Instead, that even if the critics of these systems are correct, and LLMs lack these mental capacities, their capacity to converse is unhindered.

John: Are you coming to the party tomorrow?
Jane: I have to babysit.

In this conversation, Jane does not directly answer John's question. However, given the context of the conversation, she implies that she cannot attend the party because she has other commitments, namely babysitting. Grice termed these cases implicatures (Grice, 1975). Some LLMs like GPT4 correctly identify (and respond) to implicatures without any higher-order mental states. Note that not all LLM-chatbots succeed in picking up on the implication in Jane's response at the same success rate. GPT4 demonstrates near-human level performance (84%) in deciphering when a dialogue contains an implicature compared to other LLMs (Ruis et al., 2023). According to Ruis et al., most other types of LLMs are "close to random" (40% - 68%) compared to humans (86.2% - 92% accuracy) in detecting implicatures.

Thus, we can justifiably claim LLMs like GPT4 succeed at partaking in complicated conversations while lacking higher-order mental states about Jane or her response. Whether or not LLMs have beliefs (including higher-order beliefs), intentions, desires, and so on is part of growing research (Sartori & Orrù 2023). According to Sartori and Orrù, LLMs seem to surpass neurotypical humans at many cognitive tasks but also fail in others (causal reasoning and planning). I continue to remain agnostic on this matter.

For some philosophers it is precisely these cognitive and mental capacities that LLMs and other AI systems must possess prior to considering them as agents. For example, although unconcerned with social agency, both Brey (2014) and Véliz (2021) argue that artifacts cannot be moral agents because they lack consciousness or higher-order mental states like intentions, beliefs, desires, the ability to plan, and so on. Of course, I am not worried about moral agency here, but the claims against social agency for LLMs might be similarly motivated. One of the strictest requirements comes from Himma, who demands that consciousness is necessary for agency (never mind social and moral agency) (Himma, 2009, p. 27). However, if we take the findings of Ruis et al. seriously, these mental states are unnecessary to attribute conversational capacities to LLM-agents.

Keep in mind that sociality is exceptionally complicated, with entities and agents expressing it in diverse ways. Consequently, the paper's ambitious component proposes a liberal framework designed to encompass social agency across multiple and gradated dimensions. I devise a framework equipped to integrate emerging social facts involving social AI systems – for example, facts like the growing use of sexbots

(Adashade, 2017; Fan & Cherry, 2021; Jecker, 2021) and other facts involving human-AI system relationships (Jozuka et al., 2018; Weber-Guskar, 2021; Yang, 2020).[15]

As such, in one way, I develop a theoretical framework for chatbots' social agential status. At the same time, the framework also specifies the conditions a chatbot (or mutatis mutandis for other AI systems) should satisfy to be considered a proper implementation as a social agent. This means that I begin with the social phenomenon in the world—chatbots and their use in social settings. Then, I apply existing social theories to make sense of what is happening and to determine whether chatbots rightfully belong in our social ontology as social agents. When the relevant theories cannot account for chatbots, I tweak them. The tweaking allows us to account for these new facts concerning human-robot interactions (HRIs). When engineers, designers, and programmers decide to build something, they use an opposite strategy to the one commonly used to develop theories. Raymond Turner calls this the "intentional shift." The goals of engineers, as opposed to theoreticians, shift from explanation to specification (Turner, 2020). While developing a theory, if there is a mismatch between the theory and a phenomenon or some preexisting artifact, we change the theory or explanation, not the phenomenon or artifact (e.g., the theory of evolution). In the development of an artifact, if there is a mismatch between the artifact and specification, we alter the artifact to fit the specification; we don't change the specification to align with the artifact's traits. I partake in something of a hybrid approach here. The phenomenon of social chatbots needs an explanation, but as said above, many theories of social ontology already exist.

For instance, according to threshold approaches, AI systems like chatbots do not count as social agents (with many denying them agency). This is where the intentional shift is relevant. People are and will continue to create social relationships with AI systems, like chatbots, holograms, robots, and digital avatars. For instance, consider Replika and other chatbots marketed as friend bots or therapists (Sedlakova & Trachsel, 2023). Another example of people creating social relationships with bots is the Snapchat influencer Caryn Marjorie. She created an AI-generated avatar of herself, called CarynAI (Sternlicht, 2023). It is marketed as a virtual girlfriend, which users can pay to interact with romantically (virtually).

Thus, new approaches are required to account for these trends while staying sensitive to existing tools. We can use the theories available to

15 However, I suspect the model, if successful, is useful for other scenarios too.

explain the social phenomenon of human-robot relations and the social agential status of chatbots. Next, with the help of the intentional shift, my model creates a quasi-specification for the future development of social AI systems. That is a set of criteria a system must meet to be considered a social agent.

The pre-theoretical boundaries of our social ontology have already expanded to include AI systems like Replika, Xiaoice, and sexbots. As philosophers, we have two options: (1) Deny LLMs membership in our social ontology based on theoretical grounds – this often involves claims that only organic or X-type creatures belong in our social world as social agents (threshold accounts)[16] – or (2) expand ontological boundaries concerning social agency. Humans have a history of closing our conceptual borders to adhere to dogmatic theories. Evidence of this appears in some approaches to issues such as gender identity, interracial marriages, and advocating for proper animal rights.

That said, the first option is the least appealing. It is to draw a strict line and deny them social agential status. I take the second option, developing an updated and pluralistic theory by assessing the current literature and modifying some concepts. Next, utilizing the intentional shift notion, the new theory will become the social theoretical specification for what properties an artificial agent should possess to be considered a conversational social agent. The first step in building my model is establishing the chatbot's capacity to have conversations (C) as a social property. To do so, I rely on Payton's approach (Payton, 2023, forthcoming).

2. What makes a property social?

C is a property expressed by the predicate "conversationalist," and C is a constitutively constructed property.[17] Payton agrees with Haslanger that a property is constitutively constructed "just in case social factors are 'in' the real definition of that property" (Payton, 2023, p. 6).[18] She goes on

16 To insist on going this route is to be guilty of what Dennett calls the "Cartesian trap" (see Dennett, in this volume). That is, something more than a mechanical process is required for consciousness, emotions, and so on. In this case, is to insist that something above algorithmic process is required for an entity to be a social agent capable of partaking in social acts like friendship.

17 This formulation is exactly taken from Payton (2023, p. 6). I have just substituted the 'conversationalist' for 'cool'. She makes use of the concept cool articulated by Haslanger (see, Haslanger, 2012, ch. 2).

18 Payton characterizes these approaches as essence accounts (Payton *forthcoming*). Payton says that essence accounts say "that a property is social just in case it has an

to appeal to Kit Fine's notion of constitutive *essence* to establish what it takes for a property to be socially constructed. According to Payton, the essence of an entity is composed of a set of propositions. A subset of the set essence is social factors ({{social factors}…}), and how others treat the agent is one of those social factors. To channel Payton, "[chatbot] is a C iff [chatbot] is treated in certain ways by other people" (ibid). It is a social fact that people participate in conversations with LLM-chatbots. In the future, it will likely be more common. As mentioned above, people have already formed social bonds and relationships with chatbots and other social robots. In other words, C is constitutively constructed "just in case its *real definition* makes reference to social factors" (ibid). Remember, like Payton, I am concerned with establishing C as a social property. Whether C is sufficient for social agency is defended in the next section. As I will show, both how individuals treat an agent and the agent's actual capacity to carry on a conversation are required for social agency (at least as a conversationalist). The fact of how people treat entities with C property is part of what grounds entities with C property as conversationalists.[19] For example, Payton says, "say that in order to have the property of *being charismatic* [social property], you have to be funny. For simplicity, let's say that an individual S is charismatic iff S is funny." She goes on to say, "[S is charismatic iff S is funny] is a member of the set of propositions which together define what it is to be charismatic" (Payton, 2023, p. 6). So, to be charismatic, one must be funny and capable of making others laugh. Moreover, others treat them as someone funny. In other words, being funny is a necessary and sufficient condition for being charismatic.

Similarly, to be a conversationalist, the agent should have the capacity to converse as described above(via speech, text, sign, or what have you). Also, part of what makes being charismatic a socially constructed property is to be treated in certain ways by others. In the case of chatbots, people converse with them and treat them like an agent with the capacity to converse. In sum, an agent is funny if people laugh at their jokes and quips; similarly, an agent is a conversationalist if people converse with them. It is not enough to merely treat an entity like a conversationalist to deem it as possessing C. For example, talking to a chair or an average stuffed toy does not attribute C to it. This type of interaction is what

essence which includes social factors." Payton (2023) also defends this view. See Payton (*forthcoming*) for a detailed account of the various approaches on offer for what makes a social property.

19 Payton draws attention to a much deeper connection between language use and social properties. This relation is what sets apart social properties from non-social properties, which she calls the mark of the social.

Strasser and Schwitzgebel in this volume refer to as one-sided, or the social interaction from the human's perspective is "thrown into the void" (see Strasser & Schwitzgebel in this volume). As mentioned above, there must be some reciprocity and responsiveness from the social agent. So, entities like simple teddy bears do not count. Other examples of entities that do not count are gods, angels, and imaginary friends because, again, the conversation is one-sided, or the speech-acts of the human "go into the void."[20] However, a teddy bear embedded with a sophisticated LLM designed to have conversations does count.[21]

Nonetheless, social acceptance is necessary. For example, suppose nobody accepted verbal exchanges between human agents and chatbots (even in a minimal sense) as the occurrence of conversations.[22] Instead, the human-robot interface is nothing more than an intuitive way to conduct browser searches, like "to google" something. The verbal, auditory, and so on exchanges of words with Replika, Xiaoice, ChatGpt, and CarynAI are Google searches. The program merely provides one "search result" that the program calculated as the most relevant to the user relative to the input. The program just happens to present these search results in a more approachable and accessible way to its users. In this case, we do not accept or treat chatbots as agents with the capacity for conversations. Instead, they are nothing more than search engines. Thus, simply having the capacity to converse without the social factors is not enough for the agent with C to be considered a conversationalist. In the next section, I describe why both these conditions are necessary to justify social agency for LLMs in detail.

Notice that the considered property to establish AI agents as social agents cannot be any social property. For example, *being popular* is a social property, and the Google search engine is popular. I am not talking about Google, the company, but the search engine and the act of "googling" for information.

Does this entail that the Google search engine is a social agent? No, it is not. First, the Google search engine is not an agent, even in the minimalist sense.[23] Second, the property of being popular, though a social property,

20 Of course, people who believe in gods and angels will disagree.

21 Take another example from science fiction which counts. The TransAm called Kit from the 1980's show Knight Rider.

22 Remember Bender's proposal that chatbots are nothing more than stochastic parrots (Bender et al., 2021).

23 I do not defend this assertion here. As mentioned, others have provided criteria for what is required for minimal agency. For example, take Barandarian et al.'s account for minimal agency (Barandiaran et al., 2009). According to them, an entity must satisfy three conditions – 1) Individuation, 2) interactional asymmetry, and 3) normativity. To

is the wrong kind of social property for social *agency*. The attribute or property should provide the agent with the capacity to perform actions deemed social, and being popular is not such a property.[24]

In contrast, C is a social property. In addition, users of Google do not treat, behave, or carry attitudes towards Google in ways that demonstrate the consideration of Google as a social agent. In the next section, I build on this notion and, using Epstein's framework, demonstrate that possessing C is sufficient to ground and anchor chatbots as social agents.

2.1 The grounding and anchoring of chatbots as social agents

I think that Epstein's framework can help clarify the way in which someone's being a social agent is grounded in the capacity to carry on a conversation and, further, anchored in our collective acceptance that these are grounds for being a social agent. However, it is important to clarify that I am neither concerned with nor attempting to defend Epstein's system of anchoring and grounding. It is not without controversy (Di Iorio & Herfeld, 2018; Epstein, 2019c, 2019b, 2019a; Hawley, 2019; Hindriks, 2019; Mikkola, 2019; Schaffer, 2019), also see: (Khalidi, 2015). Much of the criticism of Epstein's system focuses on whether anchoring is a helpful concept. Some argue that the grounding tool can do the same

provide a brief overview, an entity unable to differentiate itself from its environment fails to meet (1). Relatedly, to satisfy (2) it ought to be able to affect or modulate its relationship with the external environment and is not entirely dependent on environmental forces. (3) Normativity asks that the interaction with the environment must occur according to some norm(s). That is, "goals or norms according to which they [agents] are acting, providing some sort of reference condition so that that interactive modulation is carried out in relation to this condition" (ibid, p. 373). Burge also formulates similar conditions for minimal agency (Burge, 2009). For example, a bird gliding through the air meets all three, but a cloud floating through the air does not. The google search engine clearly fails to satisfy all three conditions, but a full defense of this claim is unfeasible.

24 One response here is that what if a child's teddy bear says "hi" when its paw is squeezed or as a response to the child saying "hi." Is this sufficient for social agency? Let me respond to this in parts. Is the interaction a social interaction, yes, it is, though an extremely minimal social interaction. Suppose we could quantify the level a social action is social. This minimal case would be some near non-zero number. However, is it a conversation? No. As stated, the teddy bear in this case cannot carry on the conversation beyond the simple greeting or navigate any unforeseen dialogue. Lastly, is it a social agent? Here, the answer might be unsatisfactory to some. Firstly, if one accepts that possessing social property C is necessary to be a social agent as a conversationalist, then no. Perhaps it is possible to construe it as a social agent by focusing on a different social property. However, besides saying 'hi,' the teddy bear cannot perform any other actions. Thus, the only social action it is capable of is saying 'hi,' but this is insufficient for C. Thus, provided it has no other appropriate social property, it is not a social agent. For example, under Strasser & Schwitzgebel's approach in this volume, all the social deposit by the child beyond the initial 'hi' goes into the void.

work as Epstein's anchoring is supposed to do. Overall, I do not think my argument hinges on whether Epstein's system works. The 'grounding only' framework will work just as well with the proper adjustments.

I settle on Epstein's framework because I find it intuitively appealing. He differentiates between two notions. First, the frame principle describes the conditions for the grounding facts. Second is the set of facts (anchoring) that put the frame principle in place. Thus, for me, Epstein's approach is more accessible to track than grounding only theories.

A frame principle tells us what grounds what. For example, take a dollar bill; he says, "It [frame principle] provides the metaphysical reasons that something is a dollar" (Epstein, 2015, p. 76). It is important to note that the frame principle is itself not part of the conditions on what grounds the fact; "instead, it describes how the social fact is grounded" (ibid, 77). In other words, it is a general principle that describes the grounding condition across many situations. In the case of an American dollar bill, the frame principle, according to Epstein, is:

> Principle D: For all z, the fact z is a bill printed by the Bureau of Printing and Engraving grounds the fact that z is a dollar.

Principle D provides the grounding condition: "Billy is a bill printed by the Bureau of Printing and Engraving." This grounding fact grounds the social fact that "Billy is a dollar bill." Epstein states, "When we set up conditions for some social fact to obtain, we set up the grounding conditions for that universe" (Epstein, 2015, p. 77).

The next part of the framework is anchoring. According to Epstein, anchoring provides the "glue" that holds social kinds together. Anchoring facts set up the conditions for frame principles to provide the grounding conditions for the grounding facts. In simpler words, anchor(s) put frame principles and the grounding facts in place. There are several anchoring theories, but all serve a similar purpose. As stated by Epstein, "they are theories about the "putting in place" relation that holds between a set of facts and the grounding conditions for a kind – in other words, between a set of facts and frame principle" (ibid, 81).[25]

25 Epstein provides an example of this in law. He says, "Massachusetts law (L) contains specific conditions for having the property being a first-degree murder (81). This is anchored by "Judicial decisions, jury instructions, and trial results pertaining L. The legislature enacted and the governor signed a bill pertaining to L" (94). Similarly, though not legally encoded, Frame Principle Alpha can be anchored not by law, but by social recognition, affirming attitudes and so on towards alpha.

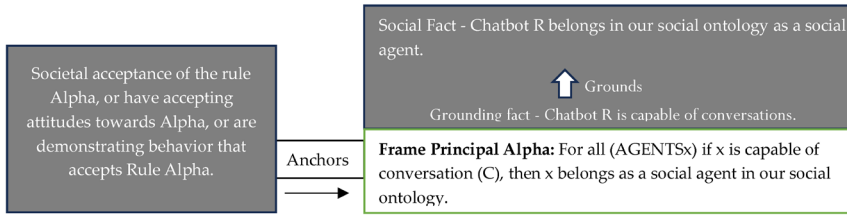

Figure 1. The grounding and anchoring of chatbots within my model of social ontology.

With Epstein's machinery in place, the rest is relatively straightforward. Let us take a chatbot called R. Figure 1 describes how a chatbot R is grounded and anchored as a social agent. Thus, we end up with the social fact – R belongs in our social ontology as a social agent. This is grounded by the grounding fact that R is capable of C (conversations) as described earlier. The grounding fact is supported by the frame principle alpha, which describes the conditions for any chatbot to satisfy. For any chatbot that satisfies the condition(s) described by alpha, then that chatbot is capable of conversations. Of course, the grounding and anchoring model takes for granted what it is for an agent to be attributed C. Some might disagree that C, as described in section two, is sufficient for being a conversationalist. Something more or different is required. In this case, the disagreement is with my formulation of what it takes to attribute C to an agent but not alpha itself. Alpha describes or specifies what it takes for an agent to be a social agent on the conversational dimension of social agency. It demands that an agent satisfy certain conditions (conversational capacities), whatever they may be. It does not describe what it takes to be a conversationalist. That is achieved by C, as described above. For me, if an agent satisfies C-conditions, then the agent is a conversationalist.

Lastly, what puts in place alpha as the description of the conditions that an agent must satisfy to be a social agent? The anchors. In the case of anchoring theories, Epstein finds Searle's (community members accept the frame principle) and Hume's (beliefs and practices towards alpha) approaches mistaken (Epstein, 2016). This is where the tweaking plays a role. In my model, users can employ any or all available theories as anchoring theories. This means that pro-attitudes, beliefs, behaviors, and acceptance of alpha can all serve as anchors.

However, someone who adheres to the threshold account might disagree with alpha, find issues with the anchoring facts, or possibly both. The latter might worry about the level of consensus required. What proportion of a community should recognize the conditions set out by alpha? I think this is likely a morally complicated question. For example,

should it be a majority of the people, should there be a referendum, or should our elected officials decide? For myself, I think, like other issues, gender, transgenderism, homosexuality, abortion, and so on, leaving it to majority rules can have dire consequences. Although these are important questions related to the motivation of this paper, they are not the focus of this paper. Thus, I leave them unanswered here.[26]

The other aspect of Figure 1 where a threshold proponent might disagree is the frame principle. The claim might go something like this: to be an agent, let alone a social agent, far more is required than simply having property C. In other words, for social agency, the agent must have beliefs, desires, self-reflective thoughts, mental states, emotions, and other properties and traits. These traits are usually associated with humans and some animals like the great apes, dolphins, and killer whales. The concern is understandable.

As I explained, sociality is a vast network of interrelated dimensions, and not all agents instantiate all the social properties. Where in our social ontology do chatbots belong, if they do at all? Contrary to the threshold approach, I argue why possessing property C is sufficient for social agency in the next section. I also explain where in the sociality network chatbots might exist.

3. Possessing a single social property is sufficient for social agency

Is a single social property (even if it is the right type) sufficient for social agency? Yes, it is. Let me explain. When we understand social agency as multi-dimensional and gradated, it allows us to include agents unable to satisfy threshold accounts. The multi-dimensional approach divides social agency into several dimensions. One of those dimensions is the conversational (or linguistic) dimension. To qualify for this dimension, an agent must satisfy C.[27] The idea is that once an agent meets certain conditions, in this case, possessing the single property C, the agent qualifies for social agency as a conversationalist. To quote Andrew Lee, "Mass comes in degree, but it is never a matter of degree whether x has mass" (Lee, 2023). The difference between my approach and threshold accounts is that what is required for social agency is far less restrictive. Yet, there are some minimal conditions. As Ruben says, "social entities, if

26 This is likely unsatisfactory to many, but to properly provide an answer is more of a distraction to the crux of this paper, than helpful.

27 It is possible that other more liberal and inclusive accounts might develop less restrictive conditions.

such there be, are entities such that they have at least one social property essentially" (Ruben, 2001, p. 14394). In the case of social agency, the social property should enable the agent to perform social actions (i.e., conversations). In short, under my model, C (or other relevant property) is sufficient, but for threshold models it is not. However, threshold models also cannot account for a variety of agents either, like animals, infants, toddlers, cognitively disabled, and so on. A pluralistic model is more appealing than a restrictive one.

The level of sophistication the agent possesses on any dimension(s) varies from agent to agent. Silver et al. also recognize that social agency, like agency, is better modeled multi-dimensionally. They separate three dimensions, "effect," "joint," and "context" (Silver et al., 2021, p. 436). In short, effect-dimension is where an action causes an effect produced by another agent. The joint dimension encompasses acts of two or more agents, "forming a joint identity. For example, when two or more people reposition a sofa" (ibid, p. 436). Lastly, the context dimension. This is a bit unintuitive. It is where social agency is produced when acting in some social context, like "being in the presence of another independent agent" (ibid, p. 436). For example, Silver et al. say, "completing the crossword of a newspaper with someone glancing over your shoulder at your progress." In this case, the action produces no effect in another (effect), nor is it done jointly (joint), but the presence of this other agent (context) influences your sense of agency. (ibid, p. 436).

Although their social agency spectrum tracks cooperation, they say, "there are many interaction dimensions critically under-researched in relation to Social Agency, and while this [their rendition] continuum is centered around the degree of cooperation in an interaction, as Social Agency grows as a field, it is hoped that more key elements will be incorporated into this model" (ibid, p. 442).

Along the same lines as Silver et al., Figure 2 provides a way to visualize the conversational dimension. A straightforward way to plot chatbots on this dimension is by focusing on the sophistication of their conversational capabilities. Another way is by only focusing on how much social agency users attribute.[28] In short, this approach only takes users' attitudes towards LLMs as the relevant factors to determine the level of social agency while ignoring the actual capacities. The better approach is to take into account both approaches (see section 2.1).

To reiterate, the sophistication of a chatbot's capacity is of little consequence if people lack beliefs or attitudes and do not accept the frame

28 Strasser & Schwitzgebel, in this volume, discuss this aspect.

principle as grounding conditions for the grounding fact. Under these circumstances, it becomes increasingly challenging to establish chatbots as social agents. Alas, that is not the case. Therefore, plotting the chatbots accurately requires both the sophistication of the conversational skills and the strength of the anchors.

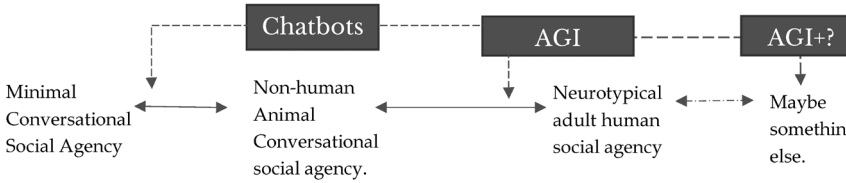

Figure 2. This scale only represents the gradated aspect of the conversational dimensions of social agency. Notice that although bacteria are considered minimal agents on some accounts, they cannot be plotted on the conversational dimension because they lack C. They might be considered social agents (likely still minimally) on a different dimension (i.e., cooperative behavior).

On one end of the scale, we have minimal linguistic agency; maybe chimpanzees who have learned sign language, rudimentary customer service chatbots, and dolphins who apparently have a localized primitive quasi-language capacity (King et al., 2013) fit those criteria. On the other end (ignoring AGI+), there are neurotypical human adults. To take an example from science fiction, the Navi people from the movie Avatar might be higher on this dimension (and others) than humans. As the movie portrays them, they can "talk" to not just each other but also other animals and the planet. Nonetheless, as far as we know, human adults seem to exhibit the most sophisticated level of linguistic agency on the conversational dimension.[29]

Anna Strasser and Eric Schwitzgebel also highlighted this point in this volume. As they and others have argued elsewhere, the point is that we must diverge away from threshold notions of agency, social agency, and likely even moral agency (Floridi & Sanders, 2004; Smids et al., 2020; Strasser,

29 There is an issue here on how exactly we quantify sociality. For instance, the coordinated actions of army ants or bees seem to be highly sophisticated and social. Does this make army ants more social than humans? Questions like these are misdirected. They operate on the assumption that sociality exists on a single dimension, that is not the case. As I have been arguing that social agency is multidimensional and different social agents occupy various dimensions and in different degrees. So, army ants might occupy higher degrees of coordination in Y-types tasks than humans, but when viewed overall and with a number of dimensions occupied by humans, humans likely possess far more social agency than army ants.

2022a).[30] However, in the next section, I further underscore why we should and how to diverge from threshold notions using an effective epistemic tool called Levels of Abstraction (LoA) (Floridi, 2008; Primiero, 2020).[31]

4. Concerns: Intentions, beliefs, and desires are necessary for social agency

The idea of levels of explanations has been around for a long time, but here, I use the concept of LoA as defined by Floridi (2008) and Primiero (2020). Abstraction helps clarify a particular phenomenon or artifact of inquiry by focusing on a set(s) of properties or details over another set(s). Usually, one set is an abstraction from some other aspect of the whole target of inquiry. Note that it does not allow one to be vague. Instead, focusing on a particular aspect of the inquiry at hand allows one to be more precise. An example from Floridi might help with the concept. He says,

> Consider the wine example. Different LoAs may be appropriate for different purposes. To evaluate a wine, the "tasting LoA," consisting of observables like those mentioned in the previous section, would be relevant. For the purpose of ordering wine, a "purchasing LoA" containing observables like *maker*, *region*, *vintage*, *supplier*, *quantity*, *price* and so on would be appropriate; but here the "tasting LoA" would be irrelevant (Floridi 2008, 309).

In this case, I divide the domain of inquiry – exercise of social agency – into two LoAs. The less abstract LoA is the macro physical world – embodied people, animals, trees, etc. In short, the physical world as we usually experience it. The second and higher LoA (more abstract) is the social, purely linguistic, or conceptual world. This is the level of language, conversations between two entities, the socio-linguistic world. In short, the way we experience the world conversationally, through language or concepts, but not physically.

	LoA	**World**	**Example**
1	more abstract	conversational	talking, chating etc.
2	less abstract/real world	The world as we usually experience it. It contains all dimensions of sociality	Going to dinners, movies etc. (can also include LoA1)

Figure 3. LoA table for social agency

30 Brey also makes a distinction between moral agency and agency (Brey, 2014).

31 LoA is similar to the idea of levels of explanation: see Floridi (2008) for an overview.

Any debate on whether chatbots or AI systems have social agential status ought to consider the relevant LoA for the system under scrutiny. For chatbots, the relevant LoA is the conversational/socio-linguistic level. Their actions take place at this level. The effect of the chatbot's actions instantiates at the LoA-conversations. For example, recall Silver et al.'s three-dimensional framework for social agency, particularly the dimension "effect." They focus on physical joint-agency and claim that joint agency between humans and robots seems questionable (though not impossible in principle). According to Silver et al., several studies have found that a joint sense of agency is not obtained with human-robot interactions, "when interacting with nonhuman agents (or robots), it has been found that implicit Social Agency enhancement disappears" (Grynszpan et al., 2019; Obhi & Hall, 2011; Sahaï et al., 2019).

In Grynszpan's study, the participants were unaware that they were interacting with a robot, but the diminished sense of agency occurred, nonetheless (Grynszpan et al., 2019).[32] That is not the case for Social Agency-effect. For example, when someone laughs at your joke, this is a social effect of your action (Silver et al., 2021, p. 436). Similarly, when we converse with a chatbot, the actions of the chatbot have social effects. Furthermore, Brandi et al. (2019) suggest that contrary to Grynszpan's conclusion that joint-agency "is human-centric," it might be possible with chatbots. In Grynszpan et al.'s study, the action in question was one of turning a knob with a partner. Two participants were sitting next to each other, separated by a curtain. Both humans were asked to turn their knobs, but both devices were connected to and effected by each other's physical state. How much resistance or force they felt when turning their knob was affected by their partner and the action and direction he or she turned his or her knob.

Moreover, sometimes the knob was turned by a robot unbeknownst to the participants. Grynszpan et al. found that "despite being unaware that their partner had changed, participants judged their contribution to the action-effect as higher when they were paired with the robot, while IB (intentional binding) occurred only with the human partner and not the robot" (Grynszpan et al., 2019, p. 7). Participants attributed to robots a lesser sense of agency than human partners, merely from the difference in the feedback. This is an interesting result for physical joint

32 Note that there was a difference in the feedback when a robot was their counterpart versus a human; they had haptic feedback, but the participant did not know which was which. They were told the whole time that it was a human on the other end. Furthermore, "diminishing Social Agency effects from interacting with robots mirror those found in human interaction" (Silver et al., 2021, p. 449).

action. However, it tells us very little about conversational joint action. It is possible that highly sophisticated chatbots, particularly those with response times more attuned to human response times, might lead to a different outcome. For example, chatbots that initiate conversations rather than merely respond to human commands might yield results contrary to those found by Grynszpan et al. My takeaway from this analysis of the psychology of how we provide agency to others, including robots, and how we measure it tells us that social agency, understood as multi-dimensional and gradated, can incorporate chatbots.

Moreover, chatbots also count under Brandi et al.'s notion of social agency, which states, "the sense of self that is gained through the perceived control one exerts over the social world. In this account, agency is not limited to one's own motor actions, but it is also implicated in social interactions and in receiving feedback from individuals" (Brandi et al., 2019, p. 18). Although chatbots lack a sense of self, they partake in social acts that exert control over the external world. They provide feedback to humans with varying types and complexity of verbal responses. Thus, chatbots affect us and our social world. Lastly, the results of Grynszpan et al. do not necessarily undermine my suggestion. They suggest that their experiment might track intentionality more so than agency (ibid, p. 9). This is consistent with my proposal. I do not propose that chatbots have full-fledged social agency (intentions, beliefs, etc.). Instead, to be a social agent requires far less than meeting these threshold notions.

Still, individuals like Brey, Veliz, Himma, Davidson, Anscombe, and others might insist on higher-order mental and psychological states as necessary. To that, let us briefly examine how sociality (amongst other human traits) developed in humans. This will show that our ancestors likely did not possess all the higher-order mental states but developed them over time.

Some theories suggest that it was not the bigger prefrontal cortex (PFC) that pushed us to be more social. Instead, it was the selective pressures forcing humans into groups for adaptive reasons that were selected for the larger PFC. PFC is largely responsible for our social capacities, including language (Preuss & Wise, 2022; Smaers et al., 2011). Thus, this supports two aspects of my argument. Firstly, and unsurprisingly, the capacity to be social evolved over a long time, coinciding with the development of the PFC and certain higher-order mental capacities (Smaers et al., 2011, p. 67). Thus, it developed in humans as a matter of degree. In other words, humans acquired dimensions of sociality via natural selection over hundreds of thousands, if not millions of years. Similar arguments are put forth for degrees of consciousness (Lee, 2023). To suggest that one is a

social agent only if one possesses these sets of higher-order mental states seems to suggest that human ancestors were not social agents, at least not until they acquired all the higher-order mental states. For example, chimpanzees and other great apes are social but are deficient in higher mental capacities. Do we want to deny them social agency? This is clearly mistaken. If cognitive evolution is any indication, then threshold notions purporting that one is not a social agent unless one possesses all the highly sophisticated mental properties are unnecessarily demanding and mistaken. For example, what if we made the same argument about other traits like flying. We consider the flight capacity of insects (that can fly) and most birds as flying. However, both of them achieve this by having very different anatomies. Interspecies and intraspecies abilities of how long, how high, and speed for both vary drastically. Suppose, as threshold proponents, we were to demand that a creature is considered capable of flight only when it possesses an x-size wingspan, hollow bones, and stays in the air for a y-length of time (maybe add speed and other traits to render it even more restrictive). All insects are out in this case, and depending on the values of x and y, many birds might be out, too. Framing the capacity to fly as such is obviously wrong. The idea is when we do not place high demands as such on other capacities, then why do so for social agency (and non-social agency, for that matter)?

Secondly, the various social capacities or properties humans possess today also developed over time for various adaptive reasons. Among those properties are language and the capacity to converse using said language (Boyd & Richerson, 2009). The capacity to converse sets us apart from other animals, including primates, in many ways. It has propelled us far ahead as social creatures, "[it] allows low-cost, generally honest communication of virtually unlimited complexity" (Boyd & Richerson 2009, p. 1). Therefore, with this powerful capacity (and others), our agential, social agential, and moral agential capacities have grown tremendously. This is challenging the all-or-nothing threshold approaches. The development of linguistic capacities over time, along with other cognitive capacities, suggests that socio-linguistic capacities did not emerge suddenly as a single, convenient mental package. It justifies unpacking the mental package notion and understanding sociality in a similar manner. That is, it is multi-dimensional, and it is a matter of degree. Different organisms exist on different dimensions and, given their cognitive capacities, enjoy varying levels of sophistication.

The advent of LLMs and other AI systems demonstrates that an agent can possess one of the properties and lack others. Nonetheless, it exists in this pluralistic social framework. This provides evidential and theoretical

grounds to justify social agency as multi-dimensional and gradated. It further underscores the framework provided here. The framework allows us to include AI systems like LLMs in our social ontology as social agents. Therefore, including LLMs in our ontology as social agents, even if they lack higher mental capacities, is entirely justified.

5. Conclusion

In this paper, I argued that LLMs implemented as chatbots belong in our social ontology as social agents. According to Payton's theories on what makes an entity social, I demonstrated that C (capacity to converse) is a social capacity. Next, I argued that because chatbots possess C, we can ground and anchor chatbots as social agents belonging in our social ontology using Epstein's framework of grounding and anchoring.

In response to potential rebuttals by threshold approaches, I also demonstrated that possessing a single social property is sufficient for social agency by arguing that social agency is multi-dimensional and a matter of degree. As a result of this analysis, the paper provided a specification for the future development of social AI systems. For a system to be considered a social agent, it must first be an agent. Second, it must possess at least one social property, allowing it to partake in sociality, like taking social actions. Not all social properties, like *being popular*, are sufficient for social agency.

Overall, the paper argues for a pluralistic framework for social agency. My model is sufficiently flexible to incorporate the growing number of social AI systems. The threshold accounts' restrictive approach is untenable for future human-robot relations. If we are to stick to these approaches, we risk alienating and stigmatizing people partaking in human-robot relations.

References

Adashade, M. (2017). Sexbot-Induced Social Change: An Economic Perspective. In J. Danaher & N. McArthur (Eds.), *Robot Sex: Social and Ethical Implications* (pp. 289–300). The MIT Press. https://doi.org/10.7551/mitpress/10718.003.0023

Barandiaran, X. E., Di Paolo, E., & Rohde, M. (2009). Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-temporality in Action. *Adaptive Behavior*, *17*(5), 367–386. https://doi.org/10.1177/1059712309343819

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922
Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1533), 3281–3288. https://doi.org/10.1098/rstb.2009.0134
Brandi, M.-L., Kaifel, D., Bolis, D., & Schilbach, L. (2019). The Interactive Self – A Review on Simulating Social Interactions to Understand the Mechanisms of Social Agency. *I-Com*, *18*(1), 17–31. https://doi.org/10.1515/icom-2018-0018
Brey, P. (2014). From Moral Agents to Moral Factors: The Structural Ethics Approach. In P. Kroes & P.-P. Verbeek (Eds.), *The Moral Status of Technical Artefacts* (Vol. 17, pp. 125–142). Springer Netherlands. https://doi.org/10.1007/978-94-007-7914-3_8
Burge, T. (2009). Primitive Agency and Natural Norms*. *Philosophy and Phenomenological Research*, *79*(2), Article 2. https://doi.org/10.1111/j.1933-1592.2009.00278.x
Calvo, P., Symons, J., & Martin, E. (2014). The Emergence of Systematicity in Minimally Cognitive Agents. In P. Calvo & J. Symons (Eds.), *The Architecture of Cognition*. The MIT Press. https://doi.org/10.7551/mitpress/9559.003.0021
Church, B. (2023, September 5). 5 types of chatbot and how to choose the right one. *IBM Blog*. https://www.ibm.com/blog/chatbot-types
Crespi, B. J. (2001). The evolution of social behavior in microorganisms. *Trends in Ecology & Evolution*, *16*(4), 178–183. https://doi.org/10.1016/S0169-5347(01)02115-2
Davidson, D. (2001). *Actions, Reasons, and Causes* (2nd ed). Clarendon Press ; Oxford University Press.
Di Iorio, F., & Herfeld, C. (2018). Book Review: *The Ant Trap: Rebuilding the Foundations of the Social Sciences*, by Brian Epstein. *Philosophy of the Social Sciences*, *48*(1), 105–128. https://doi.org/10.1177/0048393117724757
Dretske, F. I. (1999). Machines, Plants and Animals: The Origins of Agency. *Erkenntnis*, *51*(1), 523–535. https://doi.org/10.1023/A:1005541307925
Epstein, B. (2015). *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199381104.001.0001
Epstein, B. (2016). A Framework for Social Ontology. *Philosophy of the Social Sciences*, *46*(2), 147–167. https://doi.org/10.1177/0048393115613494
Epstein, B. (2019a). Anchoring versus Grounding: Reply to Schaffer. *Philosophy and Phenomenological Research*, *99*(3), 768–781. https://doi.org/10.1111/phpr.12644
Epstein, B. (2019b). Replies to Hawley, Mikkola, and Hindriks. *Inquiry*,

62(2), 230–246. https://doi.org/10.1080/0020174X.2018.1502935
Epstein, B. (2019c). Sociol Ontology. In L. Mclntyre & A. Rosenberg (Eds.), *The Routledge companion to philosophy of social science* (First issued in paperback, pp. 240–253). Routledge, Taylor & Francis Group.
Eshleman, A. (2001). *Moral Responsibility*. https://plato.stanford.edu/archives/fall2019/entries/moral-responsibility/
Fan, R., & Cherry, M. J. (Eds.). (2021). Could You Marry a Sex Robot? Shifting Sexual Norms and the Transformation of the Family. In *Sex robots: Social impact and the future of human relations*. Springer.
Floridi, L. (2008). The Method of Levels of Abstraction. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3182086
Floridi, L., & Sanders, J. W. (2004). On the Morality of Artificial Agents. *Minds and Machines*, *14*(3), 349–379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d
Grice, H. P. (1975). Logic and Conversation. In J. L. Morgan & P. Cole (Eds.), *Syntax and semantics*. Academic press, Harcourt Brace Jovanovich.
Grynszpan, O., Sahaï, A., Hamidi, N., Pacherie, E., Berberian, B., Roche, L., & Saint-Bauzel, L. (2019). The sense of agency in human-human vs human-robot joint action. *Consciousness and Cognition*, *75*, 102820. https://doi.org/10.1016/j.concog.2019.102820
Haslanger, S. (2012). Ontology and Social Construction. In *Resisting Reality: Social Construction and Social Critique* (pp. 83–112). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199892631.001.0001
Hawley, K. (2019). Comments on Brian Epstein's *The Ant Trap*. *Inquiry*, *62*(2), 217–229. https://doi.org/10.1080/0020174X.2017.1289694
Himma, K. E. (2009). Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent? *Ethics and Information Technology*, *11*(1), 19–29. https://doi.org/10.1007/s10676-008-9167-5
Hindriks, F. (2019). Epstein on groups: Virtues of the status account. *Inquiry*, *62*(2), 185–197. https://doi.org/10.1080/0020174X.2017.1289696
Hornsby, J. (2018). Acts According to Hyman. *Philosophy and Phenomenological Research*, *97*(1), 238–242. https://doi.org/10.1111/phpr.12507
Hyman, J. (2015). *Action, knowledge, and will* (First Edition). Oxford University Press.
Jecker, N. S. (2021). Social Robots for Later Life: Carebots, Friendbots and Sexbots. In R. Fan & M. J. Cherry (Eds.), *Sex robots: Social impact and the future of human relations* (pp. 20–40). Springer.
Jozuka, E., Sato, H., & Chan, A. (n.d.). *Beyond Dimesnsions: The Man Who Married a Hologram*.

Khalidi, M. A. (2015). Three Kinds of Social Kinds. *Philosophy and Phenomenological Research*, *90*(1), 96–112. https://doi.org/10.1111/phpr.12020

King, S. L., Sayigh, L. S., Wells, R. S., Fellner, W., & Janik, V. M. (2013). Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1757), 20130053. https://doi.org/10.1098/rspb.2013.0053

Lee, A. Y. (2023). Degrees of consciousness. *Noûs*, *57*(3), 553–575. https://doi.org/10.1111/nous.12421

Mecacci, G., Calvert, S. C., & Santoni De Sio, F. (2023). Human–machine coordination in mixed traffic as a problem of Meaningful Human Control. *AI & SOCIETY*, *38*(3), 1151–1166. https://doi.org/10.1007/s00146-022-01605-w

Mikkola, M. (2019). Grounding and anchoring: On the structure of Epstein's social ontology. *Inquiry*, *62*(2), 198–216. https://doi.org/10.1080/0020174X.2017.1312743

Nailer, T. (2022). *Moral Agency*. The University of Adelaide.

Nyholm, S. (2018). Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci. *Science and Engineering Ethics*, *24*(4), Article 4. https://doi.org/10.1007/s11948-017-9943-x

Nyholm, S. (2023). Technological Friends, Lovers, and Colleagues. In *This is technology ethics: An introduction* (pp. 213–238). John Wiley & Sons, Inc.

Oborny, B. (2019). The plant body as a network of semi-autonomous agents: A review. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1774), 20180371. https://doi.org/10.1098/rstb.2018.0371

Parthemore, J., & Whitby, B. (2014). Moral Agency, Moral Responsibility, and Artifacts: What Existing Artifacts Fail to Achieve (and Why), and Why They, Nevertheless, Can (and Do!) Make Moral Claims upon Us. *International Journal of Machine Consciousness*, *06*(02), 141–161. https://doi.org/10.1142/S1793843014400162

Payton, D. (2023). Searching for social properties. *Philosophy and Phenomenological Research*, *106*(3), 741–754. https://doi.org/10.1111/phpr.12901

Payton, D. (Forthcoming). Social Properties. In *Routledge Handbook of Properties*. Routledge.

Popat, R., Crusz, S. A., & Diggle, S. P. (2008). The social behaviours of bacterial pathogens. *British Medical Bulletin*, *87*(1), 63–75. https://doi.org/10.1093/bmb/ldn030

Preuss, T. M., & Wise, S. P. (2022). Evolution of prefrontal cortex. *Neuropsychopharmacology*, *47*(1), 3–19. https://doi.org/10.1038/s41386-021-01076-5
Primiero, G. (2020). *On the foundations of computing* (First edition). Oxford University Press.
Ruben, D.-H. (2001). Social Properties (Facts and Entities): Philosophical Aspects. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 14391–14397). Elsevier. https://doi.org/10.1016/B0-08-043076-7/01003-2
Schaffer, J. (2019). Anchoring as Grounding: On Epstein's the Ant Trap. *Philosophy and Phenomenological Research*, *99*(3), 749–767. https://doi.org/10.1111/phpr.12645
Schwitzgebel, E. (2020). Against the mind package view of minds: Comments on Carrie Figdor's *Pieces of mind*. *Mind & Language*, *35*(5), 671–676. https://doi.org/10.1111/mila.12288
Schwitzgebel, E. (2023, November 30). How We Will Decide that Large Language Models Have Beliefs [Blog Post]. *The Splintered Mind*. https://schwitzsplinters.blogspot.com/2023/11/how-we-will-decide-that-large-language.html
Schwitzgebel, E., & Shevlin, H. (2023, March 5). *Opinion: Is it time to start considering personhood rights for AI chatbots?* Los Angeles Times. https://www.latimes.com/opinion/story/2023-03-05/chatgpt-ai-feelings-consciousness-rights
Sedlakova, J., & Trachsel, M. (2023). Conversational Artificial Intelligence in Psychotherapy: A New Therapeutic Tool or Agent? *The American Journal of Bioethics*, *23*(5), 4–13. https://doi.org/10.1080/15265161.2022.2048739
Silver, C. A., Tatler, B. W., Chakravarthi, R., & Timmermans, B. (2021). Social Agency as a continuum. *Psychonomic Bulletin & Review*, *28*(2), 434–453. https://doi.org/10.3758/s13423-020-01845-1
Sims, R., & Yilmaz, Ö. (2023). Stigmergic coordination and minimal cognition in plants. *Adaptive Behavior*, *31*(3), 265–280. https://doi.org/10.1177/10597123221150817
Smaers, J. B., Steele, J., Case, C. R., Cowper, A., Amunts, K., & Zilles, K. (2011). Primate Prefrontal Cortex Evolution: Human Brains Are the Extreme of a Lateralized Ape Trend. *Brain, Behavior and Evolution*, *77*(2), 67–78. https://doi.org/10.1159/000323671
Smids, J., Nyholm, S., & Berkers, H. (2020). Robots in the Workplace: A Threat to—or Opportunity for—Meaningful Work? *Philosophy & Technology*, *33*(3), 503–522. https://doi.org/10.1007/s13347-019-00377-4
Sternlicht, A. (2023, May 12). *Some people are freaking out about CarynAI*. Fortune. https://fortune.com/2023/05/12/carynai-virtual-girlfriend-

death-threats-caryn-marjorie/
Stoutland, F. (2008). The Ontology of Social Agency. *Analyse & Kritik*, *30*(2), 533–551. https://doi.org/10.1515/auk-2008-0210
Strasser, A. (2018). Social Cognition and Artificial Agents. In V. C. Müller (Ed.), *Philosophy and Theory of Artificial Intelligence 2017* (pp. 106–114). Springer International Publishing. https://doi.org/10.1007/978-3-319-96448-5_12
Strasser, A. (2022a). Distributed responsibility in human–machine interactions. *AI and Ethics*, *2*(3), 523–532. https://doi.org/10.1007/s43681-021-00109-5
Strasser, A. (2022b). From Tool Use to Social Interactions. In Janina Loh, Wulf Loh (Ed.), *Social Robotics and the Good Life: The Normative Side of Forming Emotional Bonds With Robots* (pp. 77–102). transcript Verlag.
Strawson, G. (1994). The Impossibility of Moral Responsibility. *Springer*, *75*(1/2), 5–24.
Symons, J., & Abumusab, S. (2024). Social Agency for Artifacts: Chatbots and the Ethics of Artificial Intelligence. *Digital Society*, *3*(1), 2. https://doi.org/10.1007/s44206-023-00086-8
Turner, R. (2020). Computational Intention. *Studies in Logic, Grammar and Rhetoric*, *63*(1), 19–30. https://doi.org/10.2478/slgr-2020-0025
Van Hateren, J. H. (2015). The origin of agency, consciousness, and free will. *Phenomenology and the Cognitive Sciences*, *14*(4), 979–1000. https://doi.org/10.1007/s11097-014-9396-5
Van Hateren, J. H. (2022). Minimal Agents. In L. Ferrero (Ed.), *The Routledge handbook of philosophy of agency* (pp. 91–100). Routledge.
Véliz, C. (2021). Moral Zombies: Why Algorithms are not Moral Agents. *AI & SOCIETY*, *36*(2), 487–497. https://doi.org/10.1007/s00146-021-01189-x
Weber-Guskar, E. (2021). How to feel about emotionalized artificial intelligence? When robot pets, holograms, and chatbots become affective partners. *Ethics and Information Technology*, *23*(4), 601–610.
Yang, M. (2020). Painful conversations: Therapeutic chatbots and public capacities. *Communication and the Public*, *5*(1–2), 35–44. https://doi.org/10.1177/2057047320950636