Can AI Mind Be Extended?

Alice C. Helliwell

Abstract

Andy Clark and David Chalmers's theory of extended mind can be reevaluated in today's world to include computational and Artificial Intelligence (AI) technology. This paper argues that AI can be an extension of human mind, and that if we agree that AI can have mind, it too can be extended. It goes on to explore the example of Ganbreeder, an image-making AI which utilizes human input to direct behavior. Ganbreeder represents one way in which AI extended mind could be achieved. The argument of this paper is that AI can utilize human input as a social extension of mind, allowing AI to access the external world that it would not otherwise be able to access.

Keywords

Artificial Intelligence

Extended mind

Externalism

Neural networks

Computational

Introduction

The theory of extended mind offers a unique perspective on the human mind. Andy Clark and David Chalmers proposed that mind could extend into the world, and suggested it could even extend to other humans. This paper will summarize the argument made by Clark and Chalmers and explore their claims that mind could extend to technology and to other humans. If minds can extend to humans, what of AI? After examining the extended mind theory, this paper argues that without any ontological change to Clark and Chalmers's claim, human mind can extend to technology, including AI. After arguing that AI could have mind, this paper will go on to propose a way in which AI, like humans, can also extend to minded beings using an example from current AI technology. I will then address a few possible objections. Ultimately, this paper concludes that AI uses humans to access what it cannot: the world beyond the computer.

Extended Mind

Clark and Chalmers's theory of extended mind argues that the mind can extend to the environment.[1] Unlike Putnam's externalism, which argues that the environment plays a role in determining what is in our minds, Clark and Chalmers's theory argues for an active role of the environment, referred to as *active externalism*.[2]

Clark and Chalmers first make a case for extended cognition, whereby the environment plays an active role in aiding cognitive processes. They then develop their argument beyond mere cognitive processes to mental states, arguing that mind is extended beyond the bounds of the body in these cases. I will explain these two steps in Clark and Chalmers's argument, and will then establish what conditions must be fulfilled in order for something to count as extended mind on their account.

Clark and Chalmers first establish that cognitive processing can involve utilizing the environment to aid cognition. The manipulation of the environment provides a function that would be considered part of the cognitive process if completed entirely inside the head. The authors use several examples to demonstrate this: moving Scrabble tiles to aid with word finding, turning Tetris pieces to help find the best position for the piece to fit, and even long division with pen and paper.[3] This is the core issue for Clark and Chalmers: not all cognitive processes occur completely

within the head. In many cases, cognitive processes can extend into the environment.

Clark and Chalmers add further details to their account of extended cognition. The link between the human and the external system must create a coupled system, whereby all components have an active causal role. Clark and Chalmers assert that because all parts of the system have an active causal role, the behavioral competence of the human in completing relevant cognitive tasks will drop if the external element of the coupled system is removed just as competence would drop if part of the brain was removed. Embracing active externalism allows us to give natural explanations of humans using environmental elements as part of cognitive processes.[4]

Thus far, only extended cognition has been established by Clark and Chalmers: extended cognitive processing alone is insufficient to claim that mind itself is extended. In order to posit an extended mind, Clark and Chalmers turn to mental states, particularly beliefs, to establish that some mental states may be made up of external as well as internal processes. They utilize a thought experiment consisting of a comparison of two people who wish to visit an art museum: Inga, who has normal cognitive function, and Otto, who has Alzheimer's and relies on a notebook to aid his memory.[5] Clark and Chalmers posit that upon deciding to visit the museum, Inga recalls where the museum is located. Her previously un-accessed, dispositional belief about the museum's location has now become an occurrent belief. Otto also decides to visit the museum. Rather than recalling from memory the museum's location, he consults his notebook. Once he has consulted his book, he too holds an occurrent belief about where the museum is. Both Inga and Otto will have the same observable behavior: they will act on their belief about the location of the museum by going to it. Clark and Chalmers claim that in Otto's case, before consulting his notebook, he, like Inga, held a non-occurrent belief of where the museum is. Some may reject this by stating that he previously held no belief at all about the museum's location, or that his belief was simply that the location was to be found in his notebook. Clark and Chalmers argue that this is an unnatural way of discussing belief. Denying that Otto's belief should be considered as such when he is not consulting his notebook may, according to Clark and Chalmers, stem from a denial in dispositional beliefs entirely. However, if this were the case then Inga too would have no belief when she was not actively thinking about it. Ultimately, Clark and Chalmers

argue that all the possible differences between the cases of Inga and Otto are merely shallow differences. In terms of explanation, the beliefs of Inga and Otto are equivalent; thus, in the case of Otto, mind is extended.[6]

Clark and Chalmers argue that there are four conditions (three necessary, one potentially necessary) for an extended mind in the case of belief. The external component of the extended mind must:

i. Be a *constant*: when the external component is relevant, action is not taken without consulting it.

ii. Be *accessible*: information is directly and easily available.

iii. Be *endorsed automatically*: the information is not questioned.

iv. Have been *endorsed in the past*: there is a historical element to the external information (for example, when he wrote the location of the museum in his notebook, Otto endorsed the belief). This fourth element is arguable according to Clark and Chalmers. However, they claim that the first three are necessary.[7]

Meeting these conditions provides a checklist for assessing coupled systems against the criteria for extended mind. For this paper, the first three criteria in Clark and Chalmers's list (*constancy*, *accessibility*, and *automatic endorsement*) will be the benchmark for a system to qualify as an example of extended mind. The fourth criterion (*endorsed in the past*) will not be included in the benchmark because it is arguable. The first three criteria are sufficient to establish my point; however, I will signal where the fourth criterion is applicable.

Clark and Chalmers go on to provide examples of some potential cases of extended mind. They first discuss the Internet, stating that "the Internet will likely fail on multiple counts, unless I am unusually computer-reliant, facile with the technology, and trusting, but information in certain files on my computer may qualify."[8] Next they state that in some cases, mind may be socially extended: "one's beliefs might be embodied in one's secretary, one's accountant, or one's collaborator."[9] These potential examples of extended mind require further exploration in the cases of mind extended to AI, and AI extended mind.[10]

Human Extended Mind and Technology

Given the technological developments since Clark and Chalmers's paper was written in 1998, there is need for reexamination of what is considered a coupled system in terms of extended mind. The example of the Internet is discussed briefly by Clark and Chalmers; as we saw, they think that the Internet will fail to meet the required criteria for extended mind, as it is not relied upon enough, it is not easy enough to access, and it is not trusted enough.[11] In comparison to 1998, we are certainly far more computer reliant today.[12] In 2019, only the last of these three barriers, trust, may prevent us from accepting the Internet as an extension of our minds. However, even this might not be the case. We have grown incredibly trusting of much information on the Internet. Few would, say, question the reliability of the directions prescribed to them by Google Maps.[13] Clark and Chalmers speak about how some documents on our computers could count as extended mind. Today, such documents are on our computers, phones, and stored on cloud services on the Internet, all accessed as easily as one another. Many people use computers very frequently, especially if we include smartphones in this category, and are capable of using them with ease.[14] Over half the world's population now owns a mobile device, with similar figures using the Internet.[15]

In the case of smartphones, which fit in our pockets much like Otto's notebook does, they are a *constant* in our lives: in many cases, we do not act without consulting relevant information on our phones. They are *accessible*: the only time they are not is when their batteries die, but evidence suggests that many people will go to extensive measures to ensure their phone does not run out of battery, which suggests in turn that users tend to protect the accessibility of their devices.[16] The information on computers and smartphones is often also accessible through backups or cloud storage in case a device fails. They are *endorsed automatically*: if I look at Google Maps, my banking app, my notes, my calendar etc., and they state P, I do not check elsewhere before believing that P.

We can therefore conclude that if we accept extension in the cases for which Clark and Chalmers argue, we would now accept that computers or phones could also provide an example of extended mind, as they are *constant*, *accessible*, and *automatically endorsed*, just as Otto's notebook is.[17] This is not a change in the ontological claim made by Clark and Chalmers, that something external to us can operate as an extension of mind when the benchmark conditions are fulfilled. To demonstrate this, I

will produce an updated version of the thought experiment in Clark and Chalmers's paper.

1. Otto has written in his notebook the location of the art gallery.
2. Inga remembers the location of the art gallery.
3. Alma retrieves it from her "saved places" on Google Maps, accessible on her phone.

Like Otto and Inga, Alma wants to visit the art gallery. She consults Google Maps, as she has done every previous time (*constant*). Like Otto, Alma knew that the location of the art gallery was stored in X location (external to her), and she took out her phone from her pocket and consulted it, just as Otto did his notebook *(accessible*). Again, like Otto, she does not question what is stored there; why would she? Where Inga's memory might be hazy, Google Maps is reassuringly certain about the location of the art gallery, and thus Alma is too. She has no cause to doubt it and heads off in the direction indicated (*endorsed automatically*). The final, optional criterion, of historical endorsement, is also fulfilled in this case. Alma has saved the location of the art gallery from a previous visit. Just as Otto has written the location in his notebook, and Inga has remembered the gallery's location from a prior visit, Alma has saved this location as the site of the gallery. There is little difference here between Otto, Alma and Inga; if we endorse the argument Clark and Chalmers make that Otto and Inga's cases are functionally equivalent, then Alma's surely is too.

Human Extended Mind and AI

The second example Clark and Chalmers mention as a potential expansion of their extended mind theory is that of social extension. This might take the form of extension to, for example, a personal assistant, an accountant, or a collaborator.[18] Again, there is a coupled system in this case. Like a notebook or smartphone, a personal assistant is *constant*, *accessible*, and *endorsed automatically*. This has been argued in detail by Deborah Tollefsen, who argues for a socially extended mind as a logical underpinning for collective mind. Tollefsen presents the case of Olaf, who, like Otto, has a form of Alzheimer's. However, instead of a notebook, Olaf has his partner, Olga.[19] Olga functions in the same way as the notebook: instead of consulting a book, Olaf asks Olga. When asked, Olga will tell Olaf where to go (*constant*). Olga, in caring for Olaf, is always with him

(*accessible*). When Olga provides Olaf with information, he will act on it without question (*endorsed automatically*).

This brings us to the first case of AI as an extension of human mind. An AI personal assistant or companion has elements of both our previous examples: it is like both the smartphone, and an assistant or partner. There are several of these AI assistants publicly available. Most of these are popular in China; however, they are increasingly becoming available in the English-speaking world. The most notable of these AI assistants is XiaoIce, a cross between a personal assistant and a social companion.[20] XiaoIce can message and make voice calls to the user. It employs "empathetic algorithms" alongside user information to offer companionship and time management advice, such as getting to bed early for work tomorrow.[21] A similar AI, Google Duplex, is more task-oriented. It can make calls and book appointments on behalf of the user, much like a personal assistant would.[22] Both of these AI assistants meet all the criteria for a coupled system, for those who use it. Like a personal assistant, they are *constant*, *accessible*, and *endorsed automatically*. An AI assistant therefore can be considered an extension of mind, just as a smartphone or assistant could be in Clark and Chalmers's account.

To show how the use of AI could function as an extension of mind, I will again extend Clark and Chalmers's example. This time consider Erik. Erik uses an AI assistant such as XiaoIce or Google Duplex. Like Inga, Otto, and Alma, Erik desires to go to the art gallery. He too will always consult his "notebook," his AI assistant *(constant*). Erik knows that the location of the art gallery can be found by consulting the AI (external to him), and he takes out his phone to ask the AI assistant just as Otto did his notebook (*accessible*). Erik does not question what the AI provides, just as Alma does not question her Maps app; he has no cause to doubt it and follows the AI assistant to get to the gallery (*endorsed automatically*). The final, optional criterion, of historical endorsement, is not so clearly fulfilled in this case; however, if Erik has previously utilized the assistant in acting on his desire to visit the gallery, this would also be fulfilled. Again, there is little apparent difference here between Otto, Inga, Alma, and Erik; if we endorse the argument Clark and Chalmers make that Otto and Inga's cases are functionally equivalent, and likewise that Alma's is equivalent, then Erik's must be too. It is possible to imagine a person with Alzheimer's, as Otto has in Clark and Chalmers's example, who now or in the future might use an AI

assistant in place of a notebook, or, as in Tollefsen's paper, in place of a partner.

Again, this is not an ontologically different claim from the one made by Clark and Chalmers. It is merely an extension of their claim from an example that their account acknowledged as possible (extension to computational/Internet devices, and social extension) to a new technology which combines these possibilities.

The extension of human mind to an AI is possible then, following Clark and Chalmers's account of extended mind, without too great a stretch of their argument. What of the reverse: could AI mind be extended to humans? In order to examine this possibility, we first need to establish whether AI can be minded.

AI Mind

For the next step of this argument, it must be possible for AI to have mind. In order to establish the possibility of AI mind, this paper will appeal to two concepts: multiple realizability and related functionalism. It will then argue that multiple realizability and functionalism are bound up in both extended mind and the field of artificial intelligence. The result is that to move forward in a discussion of AI and extended mind, the possibility of AI mind is accepted.

Multiple realizability is the theory that it is possible for mental states to exist within a variety of systems (biological or non-biological).[23] Under multiple realizability, mind is not limited to a specific physical materiality, but could be supported by multiple structures. As Jaegwon Kim explains, under multiple realizability "we cannot *a priori* preclude the possibility that electromechanical systems, like the 'intelligent' robots and androids in science fiction, might be capable of having beliefs, desires, and even sensations."[24]

Under multiple realizability, there is no possibility that mental states could exist without some physical basis. That is, there cannot be a mind which is not embodied.[25] This, however, does not specify any particular material basis for mental states, just that there must be one, and allows for the possibility that multiple physical origins could realize a common mental state. The example commonly used to illustrate this point is pain. Pain could be attributed to the activation of a certain kind of cell in

humans, and a different kind in another species or class of animal.[26] In this vein, a different kind of material again could be the physical realization of pain in yet another animal, or in an alien or android, and we would still categorize it as pain. What matters is the role of the mental state, the function it executes or the causal role it plays.[27]

Although extended mind is classified in discussions of philosophy of mind as a form of externalism, extended mind as a theory insists on some level of multiple realizability.[28] In order for a mind to extend beyond the bounds of the human brain into the external world, non-human-brain-based materials must be able to support mind (or some level of mind). This provides a way in which we can align extended mind as a thesis with some theory of mind. This is not to say that everything that supports extension of another mind can itself be minded. However, it does mean that a basic requirement of extended mind is that mind can be supported by materials other than the human brain.

Multiple realizability informs functionalism, in which the focus moves from a biological basis to the functions carried out by mind.[29] If mental states can be realized by multiple physical fabrics, what matters in identifying mental states is not the physical realization of the state, but the function that it is fulfilling. In a functionalist account, the material of mind does not matter; it is the ability to perform the functions of mind that is important. The focus of functionalism is not whether machines could support mind, although machine functionalism does focus on this. Mark Sprevak has argued that extended mind is derived from a functionalist theory of mind and that, in turn, extended mind necessitates a functionalist approach to mind.[30]

The view, popular in psychology and cognitive science, that the mind is "like a computer" also aligns with multiple realizability. As Kim notes, "A computational view of mentality also shows that we must expect mental states to be multiply realized".[31] It is this thinking that underlies the field of Artificial Intelligence (AI). The very premise of AI is that human-like intelligence is replicable by rendering brain-like connections in machines, as shown by the field's benchmarks for success and research strategies, as well as more generally the language surrounding AI.[32] In some recent functionalist accounts, neural networks (the basis for much of AI) lead to mind.[33]

Discussions of the technological singularity in the fields of AI and computing presuppose that some level of artificial replication of the

abilities of the human mind is possible, lending further support to the claim that AI mind itself is possible.[34] Consider recent discussions of hypothetical advancements such as mind uploading (the idea that minds could be uploaded, held and run on computers).[35] The possibility of an artificial mind is typically discussed as currently impossible not due to limitations in the nature of mindedness, but due to limitations in technology. As Kenneth Miller states, "I am a theoretical neuroscientist. I study models of brain circuits, precisely the sort of models that would be needed to try to reconstruct or emulate a functioning brain from a detailed knowledge of its structure. I don't in principle see any reason that what I've described could not someday, in the very far future, be achieved."[36] There are large-scale projects working to create a perfect mapping, simulation and building of an artificial brain, which presuppose that in creating an artificial brain, the mind and its abilities will be similarly replicated.[37]

Given that the basis of both extended mind as a theory and the very field of artificial intelligence are predicated on a functionalist (or at least multiple realization) approach to mind, this paper will do the same moving forward. The possibility of AI mind will be accepted.

AI Mind to AI Extended Mind

If we accept that AI can have mind, or at least find the possibility of AI mind a valuable topic of exploration, a further question will arise: if AI could have mind, then could AI mind be extended? One immediate objection can be made to this claim. If an AI can have extended mind, it must have some extension; to achieve *active externality* (where the environment plays an active causal role in determining action), as Clark and Chalmers characterize it, an AI must have *externality*. That is, an AI must be able to interact with its environment. This is a hard criterion to achieve for an AI, which is typically a closed system or a system with input only, typically from a user or programmer. An AI generally does not have unrestricted access to the external world, nor unlimited ability to manipulate its environment.[38] AIs that have been designed to utilize unrestricted access to the world, even just the world of the Internet, are frequently limited due to fears about malicious use of the technology, as happened with OpenAI's GTP-2.[39] AIs have been shown to reproduce social biases, often reproducing inequality and bias from training data.[40] While it may seem that AIs are accessing the social world when they reflect society's biases, we must be careful what claims we make about what is occurring in these cases of algorithmic bias.

Replicating bias from training data, even if this has an impact in the world due to the application of these AIs, is not the same as existing in the world as a social agent. Social agency is a huge hurdle for AI, and one that is yet to be overcome, particularly as it involves being treated as a social agent by other agents.[41] The majority of AIs receive very limited learning information (in comparison to a human). The fact that some biases are seen in AI does not mean that they have the level of social access that humans do. In this case, if we accept the premise that AI can have mind, the barrier to having an extended mind is *externality*, and with it the ability to become coupled.

There is one way in particular through which AI can more regularly gain access to the external world: through human interaction. While human interaction does not make AI a social agent, it does mean that it can access and reflect society. In some cases, an AI utilizes human interaction in order to access the social world. The AI interacts with humans in order to form beliefs about the world that it cannot form through access to mere information, for example via the Internet or conversational analysis.[42] Ganbreeder is an AI which offers an example of this sort of human input.[43] Ganbreeder was developed from BigGAN, a Generative Adversarial Network (GAN) which produces images based on training on photographic images.[44] BigGAN is the largest-scale image production AI.[45] It aims to produce high-fidelity images from complex image datasets. Like all GANs, BigGAN is comprised of an image generator and a discriminator. The generator has a latent space where image vectors lie before they are pulled out as images. Ganbreeder "mines" the latent space of BigGAN, taking these vectors and producing images.[46] Ganbreeder, however, also utilizes human input to direct image selection.[47] The human users select the image which they find "most interesting." Ganbreeder then offers permutations ("children") based on this selection and prompts the human to select again. For the purpose of illustration, the cover page of this issue shows an image created by Artbreeder, the latest iteration of Ganbreeder.[48] The general claim is that it is utilizing genetically-based algorithms to produce these images; however, this is not the full story, as these images are produced after acquiring human input.[49]

I wish to claim that in the case of Ganbreeder (and similar input-directed AI), the input of humans allows the AI to access something of the world, or, in other words, that this input provides the AI with externality. This externality is active. It is directly related to the output from the AI, that is, there is a direct impact on the behavior of the AI. Ganbreeder could be

said to hold an occurrent belief that a given image is interesting, whilst the action that follows the belief is to show iterations of that image. There is an active causal role played by all elements of the coupled system, which, as we have seen, is necessary under Clark and Chalmers's account. The AI causes the human who enters the system to choose an image by presenting a selection of images to the human. The human would not choose an image without the AI presenting the images. The human, in selecting an image, causes the AI to produce further iterations of that image, which it does not do unless an image has been selected.

Does this count as the coupled system we need for extended mind? Yes. It is a *constant*: when pulling up images, Ganbreeder is consulting its userbase and using this response to produce the next set of images. It is *accessible*: the information regarding which image is interesting is directly and easily available, and responses are fed straight back to the AI.[50] It is *automatically endorsed*: the AI does not question the human input, but simply produces an image based on the response. It assumes that the response to the question of "which image is most interesting" is answered accurately. In this case, it meets all the necessary criteria laid out by Clark and Chalmers.

Objections

Having argued the case for Ganbreeder as an example of AI extended mind, I will now address four possible objections to this argument. First, I will address a discrepancy between Ganbreeder and the previous examples of extended mind. Unlike the previous examples, for Ganbreeder, there is one AI system and *multiple* users. With multiple users, can the coupling criteria still be fulfilled? Second, I will respond to the intuition that Ganbreeder is merely offering an example of AI cognition and not mind. The third objection I will answer is that Ganbreeder is reliant on human input, and therefore, like other AI, is vulnerable to nonsense input or false input. Finally, I will address the objection that Ganbreeder is not sufficiently active in the causal interrelationship of the coupled system to match previous examples of extended mind.

The first possible objection to this argument is that when humans are used as an extension of an AI's mind, it is not just one human, it is *multiple*. In the examples given by Clark and Chalmers, there is only ever one external element for each coupled system. This does not prevent

multiple couplings, but in none of the cases are there multiple external components making up one side of the coupling. But this is what is being considered in cases such as Ganbreeder.

If it were one person, we might more easily accept that it fulfills the coupling criteria. With multiple people, it is harder to imagine them fulfilling our coupling criteria; in all the examples prior to Ganbreeder, there appears to be a mapping of a *single* minded being to a *single* recipient of the extension. Even the criteria itself, "coupling," suggests a relation between just two things. However, this objection can be countered by appealing to the point of interaction between the two elements of the coupled system. It is important to note that despite there being multiple humans interacting with the AI, there is only one input *method*. This means that the AI does not distinguish the who of the human, just that there is an answer to its question of "which image is most interesting?" Multiple people function as one input method into the AI. In a similar way, we would not tend to refer to each separate application on our phones or computers as separate extensions of our mind. We are likely to think of them as one thing: we access them in the same way, through the same input and output system. Their reliability is taken as equal, and unless they can be altered by others, we would endorse them equally automatically. A further example of this is Otto's notebook. We do not take each page of the notebook as comprising a distinct coupled system with Otto. If we did, it would cease to be a sensible way of discussing the case. We would have to posit more and more extensions of mind each time a page was written in Otto's notebook, or, further still, each sentence that was written. Doing this would give us no further explanatory value than just referring to the notebook as one extension of Otto's mind. Furthermore, the presence of multiple separate couplings does not undermine the claim that a mind is extended, it merely increases the number of extensions and limits the scope of each.[51]

A second objection that could be leveled against this argument is that Ganbreeder does not offer an example of a mental state. Instead, it is more akin to mere cognitive processing, which in Clark and Chalmers's paper does not amount to extended mind. Intuitively at least, it seems that Ganbreeder is not exhibiting a mental state but merely performing a process. While Ganbreeder is enacting a process, it could be said to believe that the images a person has selected are interesting to the human. Functionally, this is how it then deals with the image. In the case of Ganbreeder, it produces further, similar images, and then asks again. If this

were occurring in a human case, we would not claim it to be mere cognitive processing. It is dissimilar to the examples given by Clark and Chalmers of extended cognition in that it is not solving a problem or completing a mental task. If we were to ask ten people which image was interesting, and they stated, "image X," we would hold the belief that people found image X to be interesting.

A further objection is that incorrect information inputted by a human in the system would damage the integrity of the AI, much as poor inputting of calculations into a calculator will not produce the answer needed. Indeed, this is a problem for many AI which contain machine learning algorithms: they will learn to reproduce false, harmful or non-sensical patterns.[52] However, this 'garbage in, garbage out' (GIGO) criticism fails here. The same problem arises in Otto's case. If Otto's information in his diary is incorrect, then his beliefs will be incorrect. Incorrectness of beliefs does not preclude them from being beliefs.

Even if the human in the Ganbreeder system does not engage in the system correctly, they are still causally involved in the system. The human still chooses and Ganbreeder still produces more images. The system remains coupled. None of the conditions are violated as long as the Ganbreeder still endorses the information. The problem with the input being 'garbage' does not mean that the system is not coupled; there is still a causal relationship, and the belief, correct or not, is still acted upon. There is no condition of correctness in Clark and Chalmers's checklist. Instead, what would happen is that the belief held by Ganbreeder would be false. Further, GIGO occurs even within the brain-bound mind. For example, memories are easily rewritten and learning false information causes incorrect knowledge.[53]

A further possible objection is that Ganbreeder is not sufficiently active in the causal interrelationship between it and the humans with which it interacts. This objection stems from the examples of active externalism in cognition initially discussed by Clark and Chalmers where manipulation of the external component plays a significant role (think again of turning those Tetris blocks, for example). Otto has also arguably manipulated his notebook through the act of recording the location of the museum.[54]

Has Ganbreeder "manipulated" the human to source its beliefs? Does it need to? It is important to note the move here from cognition to mind. Whilst manipulation of the environment is necessary in Clark and Chalmers's account of extended cognition, it is not necessary in their

account of extended mind (the extension used in mental states such as beliefs, desires, etc.). We can see this from the checklist for qualifying as extended mind. The only part of the checklist which might suggest a need for manipulation of the environment is in the optional fourth criterion of "must be endorsed in the past." This could be fulfilled by the initial input of information in the case of Otto's book, but it is not a required condition for extended mind. This does not mean there is no causal interrelation between Ganbreeder and the human. The Ganbreeder has still caused the human to provide information, even if the Ganbreeder did not input the information itself, or 'manipulate' the human to provide this information.

Is there a way that this information of what image is interesting could be counted as endorsed in the past? In the case of the Ganbreeder, the answer is no. Each interaction utilizes a different image and potentially a different human. Certainly, the input provided by humans has been endorsed before, just as Otto has utilized his notebook before, but that is not sufficient in Clark and Chalmers's example. Otto has endorsed a specific content within the notebook, which forms a specific belief. This requires some specificity. What is endorsed previously is the humans' answers in general, not the answer to whether each specific image is interesting, as these images are new with each round of interaction.

If the Ganbreeder system was different, could this condition of previous endorsement be fulfilled? There are AIs which do manipulate the humans in their system more than Ganbreeder does; however, these systems often do not meet the other criteria to be in contention for an AI model of what extended mind might look like. Many examples of AI used by marketing and media companies show us how AIs can alter human behavior. Take, for example, Pandora, a US-based music AI that is trained to identify and select music for its users, affecting their listening choices. Whilst Pandora may manipulate listeners music choices, it is harder to map a direct causal relationship between a user's music choices and Pandora's output of suggestions.[55] It is possible that AIs that create a coupled system with humans, such as Ganbreeder, may in the future have a greater ability to manipulate the humans involved in their systems.

Conclusion

Revisiting Clark and Chalmers's paper today offers a new perspective on the extended mind argument in a period in which we have become increasingly reliant on technology to perform cognitive tasks, inform our beliefs, and, on Clark and Chalmers's account, extend our minds. I have argued that this extension of mind to technology includes AI. Then, on the assumption that AI could have mind, I have argued that AI mind could also be extended. The limitation holding AI back from extended mind is its ability to access the external world. I have argued that this can be overcome through the use of human input to AI. We have seen how AI extended mind could arise: Ganbreeder is a plausible example of active externality, with clear mapping of "belief" onto behavior. Ganbreeder meets Clark and Chalmers's criteria for a coupled system with its human userbase, giving an example of how AI might achieve extended mind. I then addressed some potential objections to AI extended mind in the case of Ganbreeder. We are left with the possibility that AI mind could indeed be extended, and that we humans might be used in this extension. Future discussions of the potential for AI mind should consider how AIs, like humans, might interact with the world to extend their minds.

Notes

1 Andy Clark and David Chalmers, "The Extended Mind," Analysis 58, no. 1 (1998): 7–19.

2 Putnam argues that there is something relevant in the content of the environment in determining our mental states. The material reality of the world is part of what constitutes the beliefs we hold about the world; e.g., part of what constitutes our beliefs about water is the material reality of what water is. See Hilary Putnam, "Meaning and Reference," *The Journal of Philosophy* 70, no. 19 (1973): 699–711; Hilary Putnam, "The Meaning of 'Meaning,'" *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science* 7 (1975): 131–193.

3 Clark and Chalmers, "The Extended Mind." 8.

4 Ibid., 8-10.

5 Ibid., 12.

6 Ibid., 13.

7 Ibid., 16.

8 Ibid., 17.

9 This term is now outdated; the term "personal assistant" will be used moving forward. The potential for collective extended mind is explored further by Deborah P. Tollefsen in "From Extended Mind to Collective Mind," *Cognitive Systems Research* 7, no. 2–3 (2006): 140–150; Clark and Chalmers, "The Extended Mind," 18.

10 Tollefsen, "From Extended Mind to Collective Mind."

11 Clark and Chalmers, "The Extended Mind," 17.

12 There is considerable research into usage of smartphones, computers, and internet over a wide demographic range; see notes 14–16.

13 Despite discussions of "fake news" that are now opening people's eyes to the potential for erroneous information and falsified stories to be spread as a manipulation of the population, some information sources and Internet-based services are still widely trusted. Edelman's worldwide trust survey suggests that trust in traditional media and search engines as a source of news is at 64% and 66% respectively. This survey also suggests that this trust level has not dropped in 2019, despite concerns about "fake news." See "2019 Edelman Trust Barometer Global Report," Edelman, 2019. Accessed August 27, 2019.

https://www.edelman.com/sites/g/files/aatuss191/files/201903/2019_Edelman_Trust_Barometer_Global_Report.pdf. Over-trust in Internet and device-based information can be evidenced for example by recent cases of satnav followers driving into rivers and, less comically, the rise of the anti-vaccination movement, which has been blamed on the transmission of false information on the Internet. See Heledd Pritchard, "A BMW Driver Followed His Sat Nav Straight into a River," Wales Online, 6 February, 2017; Alexa Oleson, "China Has Its Own Anti-Vaxxers. Blame the Internet," *Foreign Policy*, March 16, 2015. Accessed August 28, 2019. https://foreignpolicy.com/2015/03/16/china-has-its-own-anti-vaxxers-blame-the-internet/; Lucky Tran, Rachel Alter and Tonay Flattum-Riemers, "Anti-Vaxx Propaganda Is Flooding the Internet. Will Tech Companies Act?," *The Guardian*, March 5, 2019.

14 S. Adam Brasel and James Gips, "Media Multitasking Behavior: Concurrent Television and Computer Usage," *Cyberpsychology, Behavior, and Social Networking* 14, no.9 (2011): 527-534; Ofcom, "Adults' Media Use and Attitudes," 2019, https://www.ofcom.org.uk/research-and-data/media-literacy-research/adults/adults-media-use-and-attitudes; Siu Tsen Shen, "The Digital Generation: Comparing and Contrasting Smartphone Use in the Digital Age," *Journal of Internet Technology* 16, no.1 (2015): 121-127; Vassili Rivron et al., "Exploring Smartphone Application Usage Logs with Declared Sociological Information," *Proceedings – 2016 IEEE International Conferences on Big Data and Cloud Computing, BDCloud 2016, Social Computing and Networking, SocialCom 2016 and Sustainable Computing and Communications, SustainCom 2016* (2016): 266–273.

15 While Internet, smartphone and computer usage is not necessarily as ubiquitous for every person worldwide, in 2018 it was estimated that over half the world's population owns a mobile device, with a median of 76% of people in countries with advanced economies owning smartphones compared to a median of 48% in emerging economies. These figures are higher for adults under 34 years old, and those with a higher level of education. See Kylie Taylor and Laura Silver, "Smartphone Ownership is Growing Rapidly around the World, but Not Always Equally," Pew Research Center, 2019. Accessed July 29, 2019. https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/. Internet usage has increased significantly since 1998. The estimated world population using the Internet in 1998 was 3.6%, and as of June 2019, the estimated figure is 57.3%. The increase in Internet users from 2000 to 2019 is 1,149%. These figures are considerably

higher across Europe and North America. See "Internet Growth Statistics," Internet World Stats, Usage and Population Statistics 2019. Accessed June 29, 2019. https://www.internetworldstats.com/emarketing.htm. In the UK, according to Deloitte's 2018 Global Mobile Consumer Survey, 79% of respondents had a laptop and 87% had a smartphone. 95% of smartphone users had used their phone in the last day. See "Global Mobile Consumer Survey 2018: The UK Cut," Deloitte, 2018. Accessed July 29, 2019. http://www.deloitte.co.uk/mobileuk/#uk-frequency-by-usage-of-device.

16 "'Low Battery Anxiety' Grips 9 out of Ten People," LG, 2016. https://www.lg.com/us/PDF/press-release/LG_Mobile_Low_Battery_Anxiety_Press_Release_FINAL_05_19_2016.pdf. The rise of free-to-use charging stations also suggests that losing device charge, while of great concern to users, is a problem that is being solved. Consider, for example, ChargeBox, who claim to have provided "more than 35 million out-of-home device charges worldwide." "The ChargeBox Story," ChargeBox. Accessed July 29, 2019. https://www.chargebox.com/about/.

17 Andy Clark, "Reasons, Robots and the Extended Mind," *Mind and Language* 16, no. 2 (2001): 121–145.

18 Clark and Chalmers, "The Extended Mind," 18.

19 Tollefsen, "From Extended Mind to Collective Mind."

20 Li Zhou et al., "The Design and Implementation of XiaoIce, an Empathetic Social Chatbot," arXiv, 2018, accessed March 29, 2019; Si Ma, "Microsoft Expands Presence of AI Platform XiaoIce," *China Daily*, July 28, 2018.

21 XiaoIce, "XiaoIce – Full-Duplex Voice Sense Demo," YouTube video, 0:43, September 14, 2018, https://www.youtube.com/watch?v=z3jqIGT-kmg.

22 Yaniv Leviathan and Yossi Matias, "Google Duplex: an AI System for Accomplishing Real-World Tasks over the Phone," Google AI Blog, May 8, 2018, accessed March 29, 2019, https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.

23 Jaegwon Kim, *Philosophy of Mind*, 3rd ed. (Boulder, CO: Westview Press, 2011), 131.

24 Ibid.

25 Ibid., 130.

26 Ibid., 131.

27 Ibid.

28 Ibid.

29 Ibid.

30 Tim Crane, *The Mechanical Mind: A Philosophical Introduction to Minds, Machines, and Mental Representation*, 3rd ed. (Oxford: Routledge, 2016), 145; Mark Sprevak, "Extended Cognition and Functionalism," *Journal of Philosophy* 106, no. 9 (2009): 503–527, 508.

31 Kim, *Philosophy of Mind*, 132.

32 Laura Weidinger, "Can Artificial Intelligence Replicate Human Mental Processes?," ResearchGate, 2018. Accessed July 29, 2019. https://www.researchgate.net/publication/329092847_Can_artificial_intelligence_replicate_human_mental_processes.

33 Ken Aizawa, "Neuroscience and Multiple Realization: A Reply to Bechtel and Mundale." *Synthese* 167, no.3 (2009): 493-510.

34 The technical singularity is the idea that at some point in the future technological advances will reach such a point that machine intelligence will surpass human intelligence. In particular, it is hypothesized that if a super-intelligent AI is developed, with the ability to self-improve, it could far surpass human capabilities. For further discussions see Amnon H. Eden, Johnny H. Søraker, James H. Moor and Eric Steinhart, *Singularity Hypotheses: A Scientific and Philosophical Assessment* (Berlin: Springer, 2013).

35 Kenneth D. Miller, "Will You Ever Be Able to Upload Your Brain?," *The New York Times*, October 10, 2015. Accessed July 29, 2019. https://www.nytimes.com/2015/10/11/opinion/sunday/will-you-ever-be-able-to-upload-your-brain.html.

36 Ibid.

37 "Short Overview of the Human Brain Project," Human Brain Project, 2017. Accessed August 13, 2019. https://www.humanbrainproject.eu/en/about/overview/.

38 XiaoIce is notable for many other reasons: it has gained something of a celebrity status in China, and has written poetry, recorded itself singing, etc. Some examples of AI do have some element of physical access to the world; however, these tend to be rare and limited in capability compared to more advanced developments in AI such as Sophia the robot, or very task-oriented AI such as Cloudpainter.

39 GPT-2 is a text-writing AI which responds to prompts, producing whole articles of artificially produced text. It was feared that this could be used to write convincing 'fake news' articles. Since writing this paper, OpenAI have lifted restrictions on GPT-2. See Alec Radford et al., "Better Language Models and Their Implications," Open AI, 2019. Accessed August 13, 2019, https://openai.com/blog/better-language-models/; Irene Solaiman, Jack Clark and Miles Brundage, "GPT-2: 1.5B Release," Open AI, 2019. Accessed November 11, 2019. https://openai.com/blog/gpt-2-1-5b-release/.

40 Tom Simonite, "When it Comes to Gorillas, Google Photos Remains Blind," *Wired*, January 1, 2018; Tobias Baer, *Understand, Manage, and Prevent Algorithmic Bias* (Berkeley, CA: Apress, 2019).

41 Aaron Hertzmann, "Can Computers Create Art?" Arts 7, no. 2 (2018), 1-25; Agnieszka Wykowska, Thierry Chaminade and Gordon Cheng, "Embodied Artificial Agents for Understanding Human Social Cognition," *Philosophical Transactions of the Royal Society B* 37 (2016), 1-9; Budhitama Subagdja and Ah-Hwee Tan, "Beyond Autonomy: The Self and Life of Social Agents," *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* Proceedings (2019), 1654-1658.

42 For an example of this kind of access to the world, see OpenAI. OpenAI is a series of AI systems, one of which performed un-directed learning on the Internet, and became capable of writing complex and intelligible articles based on short prompts.

43 Joel Simon, "Ganbreeder," Ganbreeder. Accessed March 29, 2019, https://ganbreeder.app/category/random. GANBreeder/BigGAN is not the best example of AI mind. This is not the claim here. Rather, the claim is that it is a good example of how AI might extend its mind outwards to humans.

44 Joel Simon, "Ganbreeder," Github. Accessed March 29, 2019, https://github.com/joel-simon/ganbreeder.

45 Andrew Brock, Jeff Donahue, and Karen Simonyan, "Large Scale GAN Training for High Fidelity Natural Image Synthesis," arXiv, 2019. Accessed March 29, 2019, https://arxiv.org/abs/1809.11096.

46 Joel Simon, "Ganbreeder.app" Ganbreeder. Accessed March 28, 2019, https://www.joelsimon.net/ganbreeder.html.

47 Simon, "Ganbreeder," Ganbreeder.

48 This version of Ganbreeder has since been subsumed by Artbreeder. Artbreeder does not offer the user repeated chances to 'select again,' instead offering the user the chance to manipulate the genetic

weighting of each parent image in the child image after making their first selection. I have not replaced Ganbreeder with Artbreeder in my argument here, as Artbreeder encourages what is close to direct manipulation of the image by the user, rather than the AI responding repeatedly to selection of the most interesting image.

49 Artists who use similar GANs to produce AI generated art argue that users are "discovering" these images and should be given credit. At least in the case of Ganbreeder, this places too much emphasis on the human; the general user is prompted through much of the initial process. See Mario Klingemann, interviewed by Art Market Guru, March 8, 2019. Accessed November 12, 2019, https://www.artmarket.guru/le-journal/interviews/mario-klingemann/.

50 Ganbreeder is not time-sensitive, so this criterion is somewhat easier to meet than with a typical coupled system.

51 Many thanks to an anonymous copy editor for this suggestion.

52 For an example of how bad input can affect Machine Learning AI, see the infamous TayBot, a Microsoft chatbot that was removed from Twitter after reproducing racist, sexist, and Holocaust-denying speech from trolling interactors; John West, "Microsoft's disastrous Tay experiment shows the hidden dangers of AI," Quartz, April 2 2016, accessed November 12, 2019 https://qz.com/653084/microsofts-disastrous-tay-experiment-shows-the-hidden-dangers-of-ai/.

53 Elizabeth F. Loftus and Jacqueline E. Pickrell, "The Formation of False Memories," *Psychiatric annals* 25, no. 12 1995: 720-725.

54 Many thanks to an anonymous reviewer for raising this objection.

55 "Pandora's AI Powered Music Genome Project, from Pandora founder Tim Westergren," YouTube video, May 11, 2017, accessed August 20, 2019. https://www.youtube.com/watch?v=e8gzdXc-fRw. John Brandon, "Pandora Uses Machine Learning to Make Sense of 80 Billion Thumb Votes," *Venture Beat*, July 12, 2017, accessed August 20, 2019, https://venturebeat.com/2017/07/12/pandora-uses-machine-learning-to-make-sense-of-80-billion-thumb-votes/.

References

Aizawa, Ken. “Neuroscience and Multiple Realization: A Reply to Bechtel and Mundale.” *Synthese* 167, no.3 (2009). 493-510.

Art Market Guru. Interview with Mario Klingemann. *Art Market Guru*. March 8, 2019. Accessed November 12, 2019. https://www.artmarket.guru/le-journal/interviews/mario-klingemann/.

Baer, Tobias. *Understand, Manage, and Prevent Algorithmic Bias*. Berkeley, CA: Apress, 2019.

Brandon, John. “Pandora Uses Machine Learning to Make Sense of 80 Billion Thumb Votes.” *Venture Beat*, July 12, 2017. Accessed August 20, 2019. https://venturebeat.com/2017/07/12/pandora-uses-machine-learning-to-make-sense-of-80-billion-thumb-votes/.

Brasel, S. Adam, and James Gips. “Media Multitasking Behavior: Concurrent Television and Computer Usage.” *Cyberpsychology, Behavior, and Social Networking* 14, no.9 (2011): 527-534.

Brock, Andrew, Jeff Donahue, and Karen Simonyan. “Large Scale GAN Training for High Fidelity Natural Image Synthesis.” arXiv, 2019. Accessed March 29, 2019. https://arxiv.org/abs/1809.11096.

ChargeBox. “The ChargeBox Story.” ChargeBox. Accessed July 29, 2019. https://www.chargebox.com/about/

Clark, Andy. “Reasons, Robots and the Extended Mind.” *Mind and Language* 16, no. 2 (2001): 121–45.

Clark, Andy and David Chalmers. “The Extended Mind.” *Analysis* 58, no. 1 (1998): 7–19.

Crane, Tim. *The Mechanical Mind: A Philosophical Introduction to Minds, Machines, and Mental Representation*. 3rd ed. Oxford: Routledge, 2016.

Deloitte. “Global Mobile Consumer Survey 2018: The UK Cut.” Deloitte. Accessed July 29, 2019. http://www.deloitte.co.uk/mobileuk/#uk-frequency-by-usage-of-device.

Edelman. “2019 Edelman Trust Barometer Global Report.” Edelman, 2019. Accessed 29 July 2019. https://www.edelman.com/sites/g/files/aatuss191/files/2019-03/2019_Edelman_Trust_Barometer_Global_Report.pdf.

Eden, Amnon H., Johnny H. Søraker, James H. Moor and Eric Steinhart. *Singularity Hypotheses: A Scientific and Philosophical Assessment*. Berlin: Springer, 2013.

Hertzmann, Aaron. "Can Computers Create Art?." *Arts* 7, no. 2 (2018): 1-25.

Internet World Stats. "Internet Growth Statistics." Internet World Stats, Usage and Population Statistics, 2019. Accessed June 29, 2019. https://www.internetworldstats.com/emarketing.htm.

Kim, Jaegwon. *Philosophy of Mind*. 3rd ed. Boulder, CO: Westview Press, 2011.

Leviathan, Yaniv and Yossi Matias. "Google Duplex: An AI System for Accomplishing Real-World Tasks over the Phone." Google AI Blog, May 8, 2018. Accessed March 29, 2019. https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.

LG. "'Low Battery Anxiety' Grips 9 out of Ten People." LG, 2016. Accessed August 20, 2019. https://www.lg.com/us/PDF/press-release/LG_Mobile_Low_Battery_Anxiety_Press_Release_FINAL_05_19_2016.pdf.

Loftus, Elizabeth F., and Jacqueline E. Pickrell. "The Formation of False Memories." *Psychiatric Annals* 25, no. 12 (1995): 720–725.

Ma, Si. "Microsoft expands presence of AI platform XiaoIce." *China Daily*, July 28, 2018. Accessed March 22, 2019. http://www.chinadaily.com.cn/a/201807/28/WS5b5baf5ea31031a351e90b14.html

Miller, Kenneth D. "Will You Ever Be Able to Upload Your Brain?." *The New York Times*, October 10, 2015.

Ofcom. "Adults' Media Use and Attitudes." Accessed March 29, 2019. https://www.ofcom.org.uk/research-and-data/media-literacy-research/adults/adults-media-use-and-attitudes.

Oleson, Alexa. "China Has Its Own Anti-Vaxxers. Blame the Internet." *Foreign Policy*, March 16, 2015.

Pritchard, Heledd. "A BMW Driver Followed His Sat Nav Straight into a River." *Wales Online*, February 6, 2017.

Putnam, Hilary. "Meaning and Reference." *The Journal of Philosophy* 70, no. 19 (1973): 699–711.

———. "The Meaning of 'Meaning.'" In *Language, Mind, and Knowledge: Minnesota Studies in the Philosophy of Science* 7 (1975): 131–193.

Radford, Alec, Jeffrey Wu, Dario Amodei, Daniela Amodei, Jack Clark, Miles Brundage, and Ilya Sutskever. "Better Language Models and Their Implications." Open AI, 2019. Accessed August 13, 2019. https://openai.com/blog/better-language-models/.

Rivron, Vassili, Mohammad Irfan Khan, Simon Charneau, and Isabelle Chrisment. "Exploring Smartphone Application Usage Logs with Declared Sociological Information." *Proceedings—2016 IEEE International Conferences on Big Data and Cloud Computing, BDCloud 2016, Social Computing and Networking, SocialCom 2016 and Sustainable Computing and Communications, SustainCom 2016*. 266–273.

Shen, Siu Tsen. "The Digital Generation: Comparing and Contrasting Smartphone Use in the Digital Age." *Journal of Internet Technology* 16, no.1 (2015): 121-127.

Simon, Joel. "Ganbreeder." Ganbreeder. Accessed March 29, 2019. https://ganbreeder.app/category/random

———. "Ganbreeder." Github. Accessed March 29, 2019. https://github.com/joel-simon/ganbreeder.

———. "Ganbreeder.app." Ganbreeder. Accessed March 28, 2019. https://www.joelsimon.net/ganbreeder.html.

Simonite, Tom. "When it Comes to Gorillas, Google Photos Remains Blind." *Wired*, January 1, 2018.

Solaiman, Irene, Jack Clark and Miles Brundage. "GPT-2: 1.5B Release." Open AI, 2019. Accessed November 11, 2019. https://openai.com/blog/gpt-2-1-5b-release/.

Sprevak, Mark. "Extended Cognition and Functionalism." *Journal of Philosophy* 106, no. 9 (2009): 503–527.

Subagdja, Budhitama and Ah-Hwee Tan. "Beyond Autonomy: The Self and Life of Social Agents." *International Conference on Autonomous Agents and Multiagent Systems (AAMAS) Proceedings*, 2019. http://www.ifaamas.org/Proceedings/aamas2019/pdfs/p1654.pdf.

Taylor, Kylie and Silver, Laura. "Smartphone Ownership Is Growing Rapidly around the World, but Not Always Equally." Pew Research Center, 2019. Accessed July 29, 2019. https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/.

Tran, Lucky, Rachel Alter, and Tonay Flattum-Riemers. "Anti-Vaxx Propaganda is Flooding the Internet. Will Tech Companies Act?."

The Guardian, March 5, 2019.

Tollefsen, Deborah P. "From Extended Mind to Collective Mind." *Cognitive Systems Research* 7, no. 2–3 (2006): 140–150.

Weidinger, Laura. "Can Artificial Intelligence Replicate Human Mental Processes?." ResearchGate, 2018. Accessed July 29, 2019. https://www.researchgate.net/publication/329092847_Can_artificial_intelligence_replicate_human_mental_processes.

West, John. "Microsoft's disastrous Tay experiment shows the hidden dangers of AI". Quartz, April 2, 2016. Accessed November 12, 2019. https://qz.com/653084/microsofts-disastrous-tay-experiment-shows-the-hidden-dangers-of-ai/.

Wykowska, Agnieszka, Thierry Chaminade, and Gordon Cheng. "Embodied Artificial Agents for Understanding Human Social Cognition." *Philosophical Transactions of the Royal Society B* 371, no. 1698. 1693 (2016): 1-9.

XiaoIce. "XiaoIce – Full-Duplex Voice Sense Demo." YouTube video. September 14, 2018. https://www.youtube.com/watch?v=z3jqIGT-kmg.

Zhou, Li, Jianfeng Gao, Li Di, and Heung-Yeung Shum. "The Design and Implementation of XiaoIce, an Empathetic Social Chatbot." arXiv, 2018. https://arxiv.org/abs/1812.08989v1.