Reclaiming control: Extended mindreading and the tracking of digital footprints

Uwe Peters
Center for Science and Thought, University of Bonn, Germany
Leverhulme Centre for the Future of Intelligence, University of Cambridge, UK
Department of Psychology, King's College London, UK

[This is a penultimate draft of a paper that is forthcoming in *Social Epistemology*.
Comments are very welcome.]

Abstract

It is well known that on the Internet, computer algorithms track our website browsing, clicks, and search history to infer our preferences, interests, and goals. The nature of this algorithmic tracking remains unclear, however. Does it involve what many cognitive scientists and philosophers call 'mindreading', i.e., an epistemic capacity to attribute mental states to people to predict, explain, or influence their actions? Here I argue that it does. This is because humans are in a particular way embedded in the process of algorithmic tracking. Specifically, if we endorse common conditions for extended cognition, then human mindreading (by website operators and users) is often literally extended into, that is, partly realized by, not merely causally coupled to, computer systems performing algorithmic tracking. The view that human mindreading extends outside the body into computers in this way has significant ethical advantages. It points to new conceptual ways to reclaim our autonomy and privacy in the face of increasing risks of computational control and online manipulation. These benefits speak in favor of endorsing the notion of extended mindreading.

Keywords: algorithmic tracking; digital footprint; mindreading; extended cognition

1. Introduction

When we go online to find information, do research, or communicate with others, we produce 'digital footprints': we leave behind data about ourselves in our browsing history, clicks, website visiting time, personal profiles, social media 'likes', and so on (Lambiotte & Kosinski, 2014). Many websites (Google, Amazon, Facebook, etc.) employ artificial intelligence (AI) systems such as machine learning algorithms to collect these data, recognize patterns in them, and infer our preferences, interests, and goals. Their aim is to tailor website content to us so as to keep us engaged (Lynch, 2019; Hinds & Joinson, 2019). Call the process in which algorithms monitor digital footprints and infer website users' preferences, interests, goals, etc. from them *algorithmic tracking*.

Researchers sometimes claim that in algorithmic tracking, computer systems make "psychological inferences", "judgements", and "assumptions on an individual's goals, interests and preferences" (Youyou et al., 2015, p. 4; Zanker et al., 2019, p. 190). The suggestion is that these "machines […] read our minds" in that they "infer or predict some information pertaining to […] the psychological constructs [i.e., mental states] of individual users, based on a sample of the subject's observable behaviour" (Burr & Cristianini, 2019, pp. 461, 465).

But do these descriptions adequately capture what is happening in algorithmic tracking? Or are they instances of researchers applying their "intentional stance" to AI, resulting in ascriptions of

psychological capacities to computers that they in fact lack (Dennett, 1987)? More specifically, does algorithmic tracking involve what philosophers and psychologists call "mindreading", an epistemic capacity to identify and attribute mental states to an agent to predict, explain, or shape the agent's action (Nichols & Stich, 2003; Apperly, 2011; Peters, 2019)?

It might seem clear that it doesn't. While there is debate on when exactly a being can be said to display mindreading (Halina, 2015), most minimally, it requires an agent with a mind to do the mindreading with. And the AI under consideration arguably doesn't meet that requirement (yet). Moreover, it is commonly assumed that for inferring and predicting mental states and so for mindreading, one needs to have at least a basic grasp of what a mental state is, which, in turn, requires being able to distinguish reality and appearance and understand that reality might be different from how an agent believes it to be (Dennett, 1978; Peters, 2021). Yet, the AI currently used for website personalization clearly can't do that either.[1]

However, in the following, I will provide reasons to believe that algorithmic tracking does indeed involve mindreading because humans are embedded in this process in a specific way that relates to Clark and Chalmers' (1998) "hypothesis of extended cognition" (HEC). HEC states that cognition might sometimes be partly realized by objects outside of one's body (Sprevak, 2020). I will argue that if we endorse common conditions for extended cognition, then computer systems that perform algorithmic tracking are often not merely epistemic tools that people use for mindreading. Rather, they literally extend, i.e., partly realize the human mindreading by website operators and website users.

My argument for this *extended mindreading view* assumes common conditions for extended cognition and HEC. Since HEC isn't uncontroversial (Adams & Aizawa, 2008), I will offer support for a particular set of conditions for extended cognition that provides a response to some familiar objections to HEC. I will, however, not make a case for HEC here. I rest content with a conditional argument for the extended mindreading view: '*If* we endorse common conditions for extended cognition then […].' If successful, this conditional argument still has important implications.

First, a recent review of the literature on HEC concludes that it is by "no means clear" that the arguments for HEC "cannot be made to work" (Sprevak, 2020, p. 6). That is, the case for HEC is alive. It should thus be interesting to see whether HEC also applies to mindreading. For even though HEC has been applied to various kinds of cognitions (beliefs, knowledge, etc., see ibid), it hasn't been related to mindreading yet. Similarly, while the idea that human cognition might extend into the Internet (Heersmink & Sutton, 2020; Schwengerer, 2021) isn't new, mindreading hasn't been considered in this context. And in the cognitive scientific research on mindreading, it is commonly at least tacitly assumed that mindreading is situated solely within an individual's head (Spaulding, 2020).

Second, even if the case for extended mindreading that I will develop below remains conditional, it helps to bring out that the notion of extended mindreading has in fact significant practical and ethical benefits that the alternative that human mindreading is merely embedded in, i.e., causally coupled to, computer algorithms lacks. This is because the notion of extended mindreading points to novel conceptual strategies to defend our autonomy and privacy in the light of increasing risks of computational control and online manipulation. There are thus

[1] But see also Rabinowitz et al. (2018).

[2] There are other recommender systems that don't use collaborative filtering; for details, see Smith and Linden (2017).

[3] For details on how algorithms are trained in the process of machine learning, see Burrell (2016).

ethical reasons for endorsing the notion of extended mindreading that add a new perspective to the debate on the societal implications of cognitive extensions into AI technology (Frischmann & Selinger, 2018; Hernandez-Orallo & Vold, 2019).

In sections 2 and 3, I distinguish two kinds of algorithmic tracking and motivate the claim that certain types of trait and preference ascriptions are proper instances of mindreading. In section 4, I build on these points to argue that algorithmic tracking involves human mindreading. In section 5, I introduce conditions for extended cognition, before, in sections 6 and 7, using them to support the extended mindreading view.

2. Two kinds of algorithmic tracking

When online companies monitor our digital footprints to form user profiles for website personalization, at least two types of algorithmic tracking might be involved. I shall call them *simple algorithmic tracking*, and *psychometric tracking*.

Recommender systems on Amazon or Netflix fall into the first category (Gomez-Uribe & Hunt, 2015). To produce suggestions related to what items a user might want to buy, watch, or read, the algorithms that are involved commonly (in 'collaborative' filtering)[2] (1) track a website user's previous purchases, ratings, items viewed, or clicking behavior, (2) build a model of the user on their basis, and (3) find a group of other website users whose purchases, ratings, or views overlap (Ricci et al., 2011). The algorithms then (4) aggregate items from these similar users, (5) delete items from the resulting list that the individual user has already bought, rated, or viewed, and (6) recommend the remaining ones (ibid).

This process can be viewed as *simple* tracking because when the algorithms link certain user profiles to recommendations (expressed with, e.g., 'You might also like […]'), they don't represent the recommended items *as* the user's preferences. There is no explicit categorization of people in terms of aspects of their psychology. The algorithms just predict that, based on their previous online behavior, particular website users are likely to display other kinds of online behaviors that the algorithms are designed to maximize (clicks, 'likes', etc.). While the algorithms themselves don't map these behavioral patterns onto anyone's psychology, we humans would do so and hence take the algorithms to track website users' mental states.

There are more sophisticated kinds of algorithmic tracking. For instance, Facebook individualizes each user's News Feed to highlight content that keeps users engaged. It does so by drawing on a wide range of data about users' Facebook 'likes', the profiles they visit most, whom they message, and what they specify about themselves on their profile (Lada et al., 2021). Some of these computations involve *psychometric* tracking, i.e., processing in which AI systems measure and explicitly attribute to website users psychological features including personality traits based on their online behavior, or other user data (Rust et al., 2021).

To illustrate, in a study by Youyou et al. (2015), Facebook 'likes' and personality profiles from >86,000 volunteers were collected. The personality profiles were obtained with surveys that captured (*inter alia*) respondents' political attitudes and the 'Big Five' personality traits (i.e., extroversion, agreeableness, openness, conscientiousness, and neuroticism). Machine learning was then used to train[3] an algorithm to find patterns in the data and accurately connect

[2] There are other recommender systems that don't use collaborative filtering; for details, see Smith and Linden (2017).

[3] For details on how algorithms are trained in the process of machine learning, see Burrell (2016).

individuals' Facebook 'likes' with their personality scores. When it was subsequently presented with (e.g., 300) 'likes' of website users that it hadn't encountered before, the algorithm could predict many of the users' psychological features (e.g., political attitudes and the Big Five) more accurately than even these individuals' spouses.

Unsurprisingly, AI systems for psychometric tracking[4] are now employed by, for instance, social media companies and consultant firms (see the Cambridge Analytica scandal) to segment website users into personality groups for tailored advertising and messaging (Hinds & Joinson, 2019). Indeed, IBM offers a program ("IBM Watson Personality Insights") to private individuals for predicting website users' "personality, needs, and values" based on social media data for highly targeted messaging.[5] How should we conceptualize what all these AI systems are doing in their psychometric or simple algorithmic tracking?

3. From preference and trait ascriptions to mindreading

It might be suggested that even if algorithmic tracking involved computer systems that ascribe preferences, interests, and personality traits to website users in much the same way as we humans do it, this would still not mean that algorithmic tracking also involves mindreading. For it isn't obvious that, even in humans, ascriptions of preferences and personality traits qualify as genuine instances of mindreading: Some researchers hold that trait attributions don't involve any mental state attributions, but just behavior reading (Malle et al., 2001; Andrews, 2012).

However, while work on mindreading has traditionally focused primarily and mostly on ascriptions of (false) beliefs (Phillips & Norby, 2019), many other mental state ascriptions are now recognized as instances of mindreading too. They include attributions of desires, perceptions, knowledge, stereotypes, and, importantly, preferences and personality traits (Spaulding, 2020).

The inclusion of preference and trait ascriptions is supported (*inter alia*) by neuroscientific evidence that, in humans, representations of traits and mental states have a shared neural basis (Thornton & Mitchell, 2018), and by considerations suggesting that they often involve implicit mental state attributions. Spaulding (2020) notes that when we ascribe to, say, an elderly woman the trait of being nurturing, we aren't merely drawing an inference about her behavior that she will, for example, buy gifts for her grandchildren. After all, one can do so resentfully. Rather, we predict that she will buy gifts because we assume she *cares* about the children, and *wants* them to see that she loves them. In such cases, our trait inferences involve mindreading, Spaulding holds, because they imply certain implicit attributions of mental states (emotions, desires, etc.).

I will endorse this view here: trait and preference ascriptions with underlying implicit mental state ascriptions are proper instances of mindreading. The notion of 'implicit mental state ascription' here differs from the one used in the psychological literature on "implicit mindreading" (Low & Perner, 2012). In that literature, implicit mindreading is an early developing, largely non-conceptual, non-verbal, and automatic capacity to understand mental states that is measured indirectly via, for instance, eye gaze tracking (ibid). I will here set this kind of cognition aside and will use the phrase 'implicit mental state ascription that underlies trait ascriptions' simply to mean 'mental state ascription that underlies or is presupposed by,

[4] For a meta-analysis of research on AI systems' accuracy in personality judgments, see Azucar et al. (2018).
[5] See https://www.ibm.com/watson/services/personality-insights/.

but isn't consciously represented in, trait ascriptions'. The above example of the elderly woman illustrates what is meant.

4. Algorithmic tracking involves mindreading

Even if preference or trait ascriptions qualify as mindreading, it might still seem clear that simple algorithmic tracking doesn't involve mindreading. As noted, it doesn't even involve a categorization of people in terms of preferences or interests. Even when it comes to psychometric tracking, one might hold that the mapping of online behavior onto, for instance, personality scores hardly amounts to a mental state ascription. However, I will now argue that algorithmic tracking is often a process in which website operators are embedded and attribute mental states to people. It should thus be construed as a process that itself involves mindreading.

Notice first that a website operator could be a single individual running a website, or a company, for instance, Facebook. Both may use computers for algorithmic tracking. In the case of Facebook, different people may monitor website traffic, and, as a team, perform cognitive tasks such as supervising, designing, and using algorithms (Hao, 2021). Since it is then the relevant Facebook team as a whole (not any individual member of it) that produces the related cognitive outputs, the team can be viewed as a "distributed cognitive system" (Giere, 2007). This view is compatible with holding that the cognitive outputs of that system are only realized in some of the heads of individual members (i.e., the group itself may not cognize). The point here is just that independently of whether the website operator is an individual or a company, in algorithmic tracking, there is a cognitive system managing algorithms.

In fact, if website operators (single individuals or groups) want to personalize their websites to users by tracking their browsing, clicks, 'likes', etc., they evidently can't do so without AI technology: The personal online data are either inaccessible, or too vast to keep track of for them. Website operators thus rely on analytics algorithms (e.g., Google Analytics) for monitoring website traffic. These algorithms often don't identify website users as particular persons but instead anonymize IP addresses.[6] Other analytics tools are available, however, that enable operators to connect IP addresses to specific persons and track their behavior across visits, devices, and platforms, combining anonymous and identified user behavior even for >90 days apart (such tools include, e.g., Heap Analytics[7] or Oribi[8]).

Importantly, while they typically function automatically, the algorithms used for website traffic monitoring themselves aren't entirely independent. Website operators monitor these systems to check whether they correctly track what website users are interested in, want to buy or view (e.g., by measuring conversions, time spent on the website, or ratings). Given their financial interests, website operators will also often optimize the algorithms in light of their performance (Nathani et al., 2020; Cooper, 2021). This might include retraining these systems on new data (Zhang et al., 2020), rewriting them, introducing new cookies[9] for more behavioral data, or combining different kinds of digital footprint data (Burke, 2007). Since website operators often don't just passively employ AI systems for algorithmic tracking but continuously "tinker" them

[6] For algorithms or website operators to ascribe mental states to website users and so to mindread them, they needn't also identify the users: ascriptions of mental states to anonymized users would still be instances of mindreading.

[7] https://developers.heap.io/docs/using-identify

[8] https://blog.oribi.io/track-individual-users-on-google-analytics/

[9] These are simple text files that website store on one's computer to track one's preferences.

to increasingly better identify aspects of website users' psychology (interests, etc., Cooper, 2021), website operators can be viewed as part[10] of the algorithmic tracking process.

There is then reason to believe that the algorithmic tracking itself also involves psychological ascriptions. For suppose you are a website operator who is using AI systems for simple algorithmic tracking to increase sales or user engagement. One day you learn that these systems recommend to a particular website user *U* a product or content *C* on the basis of *U*'s browsing data. Suppose you then assert:

> *P1*. 'Given *U*'s browsing behavior, the computer systems recommend *C* to *U*, but this doesn't mean that *U* is likely to be interested in, feel positive about, and want *C*.'

This assertion seems odd. If you don't assume that the AI systems capture *U*'s interests, feelings, and desires, it is hard to see why you employ them to increase sales or improve website users' online experience to begin with. The apparent oddness of *P1* is tied to the negation it involves. When we remove it, the assertion is fine:

> *P2*. 'Given *U*'s browsing behavior, the computer systems recommend *C* to *U*, meaning that *U* is likely to be interested in, feel positive about, and want *C*.'

To the extent that *P2* is a plausible claim, there is ground to believe that algorithmic tracking involves at least implicit ascriptions of mental states (interests, feelings, desires, etc.) to website users by the website operators that employ algorithms for online personalization. And since implicit ascriptions of such mental features plausibly count as instance of mindreading (Spaulding, 2020), this means that algorithmic tracking (construed as a process in which website operators are embedded) involves mindreading.[11]

5. Conditions for extended mindreading

In the remainder, I will provide reasons to believe that in algorithmic tracking, human agents aren't merely causally coupled to algorithms to perform mindreading, but the human mindreading involved is literally partly realized by these systems. This view relates to Clark and Chalmers' (1998) HEC. It will thus be useful to first introduce some motivations for HEC in general and specify conditions for extended cognition that we can then apply to algorithmic tracking.

To support HEC, Clark and Chalmers' main argument involves a thought experiment with two protagonists: Otto, who suffers from Alzheimer's, and Inga, who is healthy. To support his daily functioning, Otto routinely uses a notebook to write down and remember information. One day, Inga and Otto hear of an exhibition at an art museum and decide to see it. While Inga recalls the museum's location from her biological memory and then goes there, Otto looks up the location in his notebook and then heads in the same direction.

[10] Other human agents, companies, or networks (e.g., when the personalization algorithms are provided by third-parties; Amazon Personalize, etc.) might also be involved in this process. I will focus here primarily only on website operators.

[11] In these cases, website operators don't necessarily merely discover a user's existing, fixed, and durable mental states (e.g., preferences). Their algorithms might often shape or even create website users' preferences (e.g., via reinforcement procedures) to maximize revenues (Crockett, 2017; Frischmann & Selinger, 2018). This view is compatible with, and becomes more plausible in light of, the point here that algorithmic tracking *also* involves at least some detection of website users' existing mental states. After all, if they are able to identify users' current preferences, etc., website operators can more effectively influence and better tailor contents to them.

Clark and Chalmers argue that since Otto's notebook is functionally equivalent to Inga's belief about the museum in that it guides his museum-related action, and offers "sameness of opportunity" (Clark, 2008, p. 8), there is reason to assume that it partly realizes his belief about the museum. This is because "[w]hat makes some information count as a belief is the role it plays, and there is no reason why the relevant role can be played only from inside the body", Clark and Chalmers (1998, p. 14) hold. They hence propose a

> (1) parity principle: "If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is [...] part of the cognitive process" (ibid, p. 8).

Since this condition might result in implausibly wide extensions of the mind into, for instance, public libraries or the yellow pages (resulting in "cognitive bloat", Allen-Hermanson, 2013), Clark and Chalmers introduced additional criteria for an artefact to count as part of the mind. These conditions include that for an agent (individual or group) *S* a candidate artefact *A* also needs to be

> (2) reliably available and typically invoked (for a particular cognitive task), and
> (3) any information retrieved from *A* must be automatically endorsed and
> (4) easily accessible to *S* (Clark, 2010, p. 46).

However, Adams and Aizawa (2008) object that while these conditions might ensure a close *causal* coupling between *A* and *S*, it still doesn't follow that *A* is also *constitutive* of *S*'s cognitive system – assuming otherwise is committing a "coupling-constitution fallacy". In response, Clark and others have proposed further arguments for HEC (*inter alia*) by drawing on considerations from dynamical systems theory, according to which two systems create one overall extended system if they are in continuous bidirectional interactions (Clark, 2008, p. 80, 131).

Palermos (2014) offers the following arguments for this view. When two individual systems interact in ongoing bidirectional feedback loops to achieve a cognitive task, this results in new *systemic* properties, properties that can't adequately be attributed to one of the interactants alone. Hence, we need to postulate an extended system. Moreover, in the interactions between the two subsystems, we can't "decompose systems in terms of distinct inputs and outputs from the one to the other", because the "effects of each component to the other are not entirely endogenous to [i.e., solely originating from within] the affecting component, and *vice versa*" (ibid, p. 33). Since that is so, the two subsystems can't plausibly be viewed as merely causally coupled but are rather constitutively linked in the broader, extended system of which they are part.

To illustrate, consider (Palermo's example of) telescopic observations in which an astronomer needs to reposition the telescope while adjusting its lens, causing particular effects on the lens such that shapes appear. These shapes aren't fully specific representations of, say, planets or comets yet. Rather, what is depicted on the lens depends on the astronomer's further telescope adjustments, until planets, comets, or stars become visible. The final outcome can't be attributed to either the astronomer or the telescope alone, as neither can produce it alone. We thus need to assume that it is a property of a broader system comprising both.

Additionally, the way each of the two parts is influenced by the other isn't wholly exogenous to, i.e., solely originating from outside the other: The effects of the telescope (e.g., shapes on

its lens) on the astronomer partly originate from the astronomer's own effects on the telescope (e.g., her lens adjustments). The telescope's effects thus can't adequately be conceptualized as mere *inputs* to (i.e., as originating wholly from outside) the astronomer. And the astronomer's effects on the telescope, in turn, can't be adequately viewed as mere *output* to the telescope because they partly originate from 'within' the telescope (from how blurry its images are). Since we can't clearly separate inputs from outputs between the interactants, there is reason to postulate a system in which the astronomer's cognitive processing (e.g., object recognition) is partly realized by the telescope.

Taken together, these points support a (here) final condition for extended cognition:

> (5) *A* counts as part of *S*'s cognitive system if there is a continuous bidirectional interaction (feedback loops) between *A* and *S* (Clark, 2008, pp. 80, 131l; Palermos, 2014).

Condition (5) helps address the 'cognitive bloat' objection to HEC: Individuals aren't in continuous feedback loops with public libraries, or the yellow pages. And given (5), the postulation of extended systems no longer needs to involve a 'coupling-constitution fallacy': The emergence of systemic properties from *S-A* couplings that involve effects that can't be neatly decomposed into inputs and outputs provides a basis for the constitution claim.

Indeed, many philosophers now explicitly endorse (5) as a *sufficient* condition for extended cognition. Some hold, for instance, that "when an agent is functionally integrated through ongoing feedback loops with her social environment, the environment doesn't just causally influence her but becomes part of her character, for good or ill" (Alfano & Skorburg, 2017: 468; for other philosophers committed to similar views, see Carter & Palermos, 2016; Carter et al., 2017; Palermos, 2014, 2016; Palermos & Tollefsen, 2018). While not all advocates of HEC thus maintain that all of conditions (1)-(5) are necessary for cognitive extensions, I will here use all five as guides for determining whether a particular process qualifies as extended cognition.

6. Extended mindreading

We now have five conditions to explore whether, in algorithmic tracking, website operators or users employ AI systems merely as tools for mindreading, or whether their mindreading is literally extended into them. I will first consider interactions between website operators and algorithms before turning to website users.

6.1 Extended mindreading of website operators

To see whether website operators' interactions with computers during algorithmic tracking ever meet condition (1), i.e., the parity principle, it is useful to first reflect on, for instance, a book recommendation service that doesn't involve the Internet. Suppose you own a bookstore. Even if you don't have a computer, you could still make notes of your customers' previous purchases, ratings, store browsing, demographics, and so on. Additionally, you could segment your costumers into groups with similar profiles so that when you meet a new costumer and learn that she fits into one of the groups, you can infer and predict that, given the behavior and preferences of other members of that group, she too is likely to find certain books interesting, and may want to read them. In your thinking, you move from an interpretation of the individual's behavior to ascriptions of and inferences about her interests, feelings, and desires,

i.e., her mental states. What you are doing can thus plausibly be viewed as an instance of mindreading.

Now, many operators of online bookstores (e.g., Amazon, or individuals running commercial websites) perform the same kind of process just envisaged, but with AI systems. These systems work largely automatically. But since they are used to increase sales and improve people's online experience, website operators, including the programmers they employ, will still regularly check whether these systems do what they are designed to do (i.e., identify what people are interested in) and update them if need be. Furthermore, even if the systems produce recommendations for website users while the operators are engaged elsewhere, if the website operators were afterwards asked to motivate the systems' recommendations to a user they would arguably still mention the user's mental states (preferences, interests, etc.). After all, the identification of people's preferences, interests, and goals is the reason why algorithmic tracking is employed on websites in the first place. There is thus ground to hold that the process includes at least implicit ascriptions of mental states.

Given this and the fact that we would view the inferences involved in your case above as instances of mindreading if they happened in the head, there is reason to believe that when those inferences occur in the interactions between website operators and computer systems, they meet condition (1) for extended mindreading.

Such interactions between human agents and artifacts also often meet conditions (2)-(4). For instance, Amazon offers its algorithmic tracking systems to any company or individual interested in using them on their websites (see "Amazon Personalize") for "real-time personalized recommendations – no ML [Machine learning] expertise required".[12] Thus, in line with (2), these systems are readily available with a computer with Internet access (e.g., a smartphone). Additionally, in line with conditions (3) and (4), many website operators (individuals, Amazon, or Facebook) do typically employ AI systems for algorithmic tracking and automatically endorse their outputs, which are easily accessible to them online. In many cases, the interactions between individuals, groups, or companies and computers that are involved in algorithmic tracking will thus satisfy conditions (2)-(4).

Turning finally to condition (5), are there continuous feedback loops between *A* and *S* in the cases at hand? As noted, some website operators continuously update their algorithms in light of their performance so as to achieve increasingly faster and "more accurate" recommendations and "prediction models" to ensure market success (Zanker et al., 2019: 160; Nathani et al., 2020). To the extent that website operators monitor and optimize the algorithms they use, algorithmic tracking will involve a continuous feedback loop between humans or groups (including Internet companies) and algorithms: Website operators (1) measure the algorithms' outputs, (2) check them against user data (e.g., from users' browsing or social media) for validity, (3) update them (if they under-perform) to better track and respond to users' interests, (4) check the outputs of the altered algorithms, readjust, and so on. In this feedback loop, the cognitive feat of identifying (influencing and satisfying) website users' preferences, interests, and goals is gradually improved.

Importantly, the result can't be adequately attributed to the website operator or the computer systems alone. Neither one can do the algorithmic tracking successfully wholly on their own: Without the optimizing interventions of the website operator on these systems, the latter's

[12] https://aws.amazon.com/personalize/

outputs are likely to become misaligned with website users' preferences over time because users' profiles change. Similarly, without these systems, website operators, in turn, can't effectively track users' digital footprints. The here relevant type of algorithmic tracking thus emerges as a systemic property, as its origin can't be explained without postulating a broader system comprising both website operator and computer technology.

It might be suggested that in this broader system, website operators and computers are still only causally coupled. However, for this to be the case, we would need to be able to clearly separate inputs and outputs relative to both interactants, and there is reason to hold that this can't always be easily done. For notice that while the AI involved might program itself (Burrell, 2016), the effects of the resulting algorithms (e.g., the detection of website users' preferences) on the website operator still partly originate from the operator's own effects on these systems, namely the operator's adjustments of the parameters within which the algorithms operate (and train themselves). These systems' effects thus can't adequately be conceptualized as originating entirely from outside the website operator.

Furthermore, since the effects of the operator on the AI systems, in turn, are directly influenced by the performance of those systems, these effects too can't adequately be viewed as originating solely from (being mere outputs of) the website operator. Given the dynamic interdependence of the interactants, inputs and outputs relative to each interactant can't be clearly demarcated. This undercuts the idea that the interactants are just causally coupled and suggests that in algorithmic tracking, they instead form one single system meeting condition (5) for extended mindreading.

But what if the website operator is not an individual but a company, for instance, Facebook? Several researchers have argued that "groups of people can manifest cognitive capacities that go beyond the simple aggregation of the cognitive capacities of their individual members" such that the groups themselves count as cognizers (Theiner et al., 2010, p. 378). In fact, the arguments that support condition (5) have been used to maintain that if the members of a social group collaboratively perform a cognitive task by interacting continuously and reciprocally with each other, then they realize a socially distributed group cognition; i.e., the group then itself (vs. merely socially embedded individuals) cognizes (Gallagher, 2013; Palermos, 2016; Palermos & Tollefsen, 2018). For, as before, systemic properties will arise that can only be adequately ascribed to the whole group, and in the interactions between the group's members, inputs and outputs between them can't be clearly separated.

Now, Facebook has teams of programmers, managers, and AI researchers that do form dynamic feedback loops with each other and together monitor and control the systems involved in algorithmic tracking and website personalization (Oremus, 2016; Hardesty, 2017; Hao, 2021). Indeed, it is hard to see how any big Internet company that both uses and itself continuously tweaks its own algorithms to refine personalization outputs and maximize profit could manage such a complex task without the relevant individual employees of the company continuously mutually influencing each other in that task. It is thus plausible that if we assume condition (5), then Facebook (or a specific team within it), too, may in the here relevant cases realize group cognition. The preceding points concerning conditions (1)-(5) then equally hold for Internet companies such as Facebook and not only for individual human agents that operate websites with personalization algorithms.

6.2 Extended mindreading of website users

There is another kind of algorithmic tracking in which interactions between humans and computers meet conditions (1)-(5). It involves website *users'* mindreading, and relates to cases in which people browse the Internet for shopping or entertainment purposes without yet knowing what exactly they want. It will be useful to first consider again an example in which no computer is involved.

Suppose you are out for dinner, look at a restaurant menu, and wonder 'What do I want to eat?' You reflect on your previous choices at similar places, visualize some of the dishes on the menu, and, upon gauging your emotional responses to what you are thinking about, conclude that you would probably like to eat dish *D*. In your deliberation, you worked out what you want by running through possible dishes in your head in mental imagery and tracking your responses to what you are imagining. While this continuous feedback process might sometimes just be used for making a choice, we often also draw on it to learn about what we want. In fact, a wealth of psychological studies found that people frequently unknowingly confabulate mental states (decisions, desires, etc.) *post hoc* by interpreting their own behavior (including mental imagery; Carruthers, 2011; Cushman, 2019). These data suggest that we often lack introspective access to our own propositional attitudes, including desires, and depend on indirect, interpretive methods to learn about them (Carruthers, 2011). Think, for instance, of the way in which one might learn that one fancies a person by observing one's nervousness when being around them. It thus seems plausible enough that the just outlined kind of imagistic feedback process, too, is at least sometimes a way in which we come to know our own mental states. If that is so, then that process counts in these cases as mindreading directed at oneself: One identifies one's own mental states (desires) via gauging and interpreting one's responses to imagery.

Returning to algorithmic tracking, when using Amazon, Netflix, and other websites, people are often in a similar situation as you are in the scenario just envisaged. They visit these websites unsure of what exactly they want to buy or watch, and then outsource to AI systems the process of identifying what they want by providing websites either directly (e.g., via product ratings) or indirectly (e.g., via their browsing, past purchases, previously watched movies) with digital footprint data. The websites' AI systems then use these data, form a user profile, identify the corresponding group of similar subjects, and match preference profiles of group members with that of the individual before presenting non-overlapping choices as recommendations to the user. The user, in turn, responds with clicks, producing data that the AI systems recycle to update their offers, gradually honing in on what the user is interested in or wants to buy.

In these ongoing interactions, neither the AI nor the website user know upfront what the user wants. Both learn about (and shape) it in their interactions. And while these interactions may commonly serve users to make a choice, they also at the same time often provide users with insights into their own mental states. For instance, when a website user finally settles for a particular recommended item, she will generally take that item to be something that she indeed *wants* rather than something that she is merely disposed to click on. For she relied on algorithmic tracking to answer the question 'What do I want to watch (buy, etc.)?' and it would be strange if, when clicking on, say, 'confirm' to finalize a choice, she thought that the item is not something that she wants. This suggests that algorithmic tracking in such interactions between website users and AI often involves at least an implicit ascription of a mental state (desire) on the part of the website users to themselves. These cases can thus be viewed as situations in which website users employ algorithms for reading their own mind.

Notice that in these cases the ongoing feedback loop and mutual causation between user and algorithm functionally mirrors a process that we would in the dinner choice example above readily treat as an instance of mindreading if it occurred in the head. This means that we now have grounds to hold that these interactions between websites users and algorithms meet criterion (1) for extended mindreading.[13]

It also satisfies conditions (2)-(4). The reason is that the AI systems involved in the feedback loop are reliably available online and typically used by many people for online shopping or entertainment purposes (e.g., think of people's daily use of their personalized Netflix site). Website users also generally trust them, and they are designed and maintained for the purpose of working out (influencing and responding to) what a user wants. Additionally, they are easily accessible online with a laptop, or smartphone. Conditions (2)-(4) are thus met.

Finally, turning to condition (5), in the kind of interaction mentioned, as noted, neither the computer systems nor the website user know at the beginning what the user wants, but learn about it in their interaction. The algorithmic tracking is thus a systemic property of that interaction that can't be ascribed to only one of the two interacts alone. Neither one can produce it alone. Moreover, neither the effects of the algorithms on the website user nor the effects of the latter on the former are wholly external or internal to each other: What a user sees as a recommended item depends on (is partly determined by) the user's (indirect) impact on the algorithms (via her online behavior). It is thus partly an effect of the user herself, not merely input she receives from the computer systems.

Similarly, the website user's response to these systems in turn depends on (is partly predetermined by) the algorithms 'inner' working (and the website operators' goals). For depending on how accurately they track the user's preferences and map them onto matching recommendations, the user will respond in one way rather than another. The effects of the user on the algorithms thus aren't originating entirely from outside the algorithms either. Since the inputs and outputs in these interactions can't be neatly separated but such a separation is needed in order to hold that website users and computer systems are merely causally coupled to each other, there is a basis for holding that there is an extended system comprising both that realizes (self-related) mindreading as a systemic property.

7. Implications

While I have now provided several reasons to believe that both website operators' and users' mindreading is in some cases extended into algorithms, one might still think that the difference between this extended mindreading view and the alternative view that human mindreading is only closely causally coupled to algorithms is merely a terminological one with little practical value. I will now argue otherwise. The case for extended mindreading has significant benefits for the ethics of AI. These benefits speak in favor of endorsing the notion of extended (vs. causally embedded) mindreading.

7.1 Focusing on website users

If website users' mindreading extends into computer algorithms, then these algorithms also acquire a moral status. To illustrate, consider people with Alzheimer who may structure their

[13] Settling what one wants is commonly also an instance of practical reasoning to make a decision. Since this reasoning is part of the kind of mindreading at issue here, the argument for extended mindreading here may equally support the notion of extended decision making.

immediate living environment (e.g., by labelling objects in their flats) so that it compensates their increasing cognitive impairment by offering reminders of information. If we change such people's personalized, self-structured settings, we would reduce their memory abilities, raising clear moral problems (Drayson & Clark, 2020). Similarly, Heersmink (2017) argues that the more people depend on artifacts for their cognitive functioning, the deeper these objects are integrated with their cognitive system and so the higher the objects' moral status. Since the effects that the artifacts "have are typically stronger when the dependency is greater and the integration denser," it "matters whether cognition is merely embedded or genuinely distributed for the moral status of cognitive artifacts" (ibid, p. 28). Relatedly, if website users' mindreading is sometimes extended and not merely embedded into algorithms online, then these algorithms have a higher moral status.

This point can be taken further. As Carter and Palermos (2016, p. 549) highlight, according to "current legal and ethical theorizing", "intentional harm to a part of a person responsible for the person's mental and other faculties constitutes personal assault" when it happens without the person's consent (and no overriding considerations exist; e.g., the protection of others). Based on this point, Carter and Palermos argue that since our mental faculties can be partly constituted by external artifacts, having our integrated epistemic artifacts intentionally compromised plausibly counts as a case of personal assault, which is an action that people are generally legally protected against.

Now, in the cases when website users' mindreading extends into personalization algorithms, the latter are part of the users' minds. Hence, if we extrapolate from Carter and Palermos' rationale, then altering these systems without the users' consent could count as personal assault too. This matters because commercial websites often adjust their algorithms to maximize conversions, which might result in outputs (recommendations) that aren't in the best (e.g., financial) interest of website users. These algorithms thus aren't neutral in that the humans whose minds extend into them remain fully autonomous decision-makers – by adjusting these systems, website operators can alter people's preferences to maximize their own profit (Frischmann & Selinger, 2018). Adopting the extended mindreading view helps here because such alterations could then in some cases be construed as intentional harm to parts of website users' mind without their consent. This notion, which is unavailable on the embedded mindreading view, has thus the benefit of potentially enabling website users to refer to the mentioned ethical and legal frameworks for personal assault to protect themselves against such interventions.

Notice that in treating the extended cognitive system that comprises the personalization algorithms as partly the website *user's*, the extended mindreading view does not ignore that other people currently own and manage the algorithmic components of that system. The point is instead that it is more plausible that ownership and control over the algorithms should shift more to the website users if they interact with the algorithms so intimately that the algorithms can be viewed as part of their minds than if the interactions and loops result merely in a causal coupling. Indeed, while there are different notions of private ownership, according to Locke's (1967; pp. 285) labor theory of property, close engagement with an object (e.g., farming the land) can *create* property rights: One's work enters into the object, turning the latter into one's property (Gallagher, 2013). Given this, in highlighting the high degree of interdependence and mutual investment between website users and algorithms, the notion of extended mindreading might help redistribute the power over these systems.

The extended mindreading view can also boost people's attention to the risks of online control and manipulation connected to algorithmic tracking. For it seems plausible that the more one thinks of something as literally part of oneself, the more one cares about what others do to it. In suggesting that personalization algorithms that many of us routinely and extensively interact with online are literally parts of our minds, the extended mindreading view can help raise people's awareness of potentially harmful consequences of their Internet use (e.g., online manipulability).

This view can also contribute to the protection of people's online privacy. For Lynch (2013), privacy of thought involves that individuals can access their own thoughts in a way another person can't (i.e., 'from the inside'), and that "you can, at least sometimes, control what I know about your thoughts": you "can refrain from telling me the extent of your views and your feelings" (p. 3). If people's mindreading literally extends into computer algorithms, then website operators' access to these algorithms may represent intrusions upon website users' privacy of thought: in having one's mind open to view to website operators, the possibility for free deliberation is reduced, because one is less at liberty to privately mull things over (Fritz & Reiner, 2016). To the extent that people "tag Internet-connected algorithmic devices as their extended minds, this raises the bar for privacy, qualifying the information within the amalgam of our brains and our devices as private thoughts" (Reiner & Nagle, 2017). And if personalization algorithms are constitutive parts of website users' mindreading, this provides privacy campaigners with argumentative support to call for more protection of website users from algorithms. The extended mindreading view thus offers conceptual ways to reclaim our autonomy in the face of new forms of computational control and privacy intrusions.

7.2 Algorithms as cognitive extensions of website operators

The notion of extended mindreading also has significant ethical benefits (that the embedded mindreading view lacks) when it comes to website *operators'* mindreading via AI algorithms. To see this, notice first that one important feature that distinguishes many AI algorithms is their ability to learn (Burrell, 2016): They don't just implement human-designed rules, but create their own, for instance, by revising the rules originally designed by human programmers, or starting from scratch. Given this, if such an algorithm designed by computers makes a mistake, whose responsibility is it?

The problem is that algorithms don't fit well into current views of liability: "Liability requires injurious acts, but what does it mean for an algorithm to act? Only people act; and algorithms are not people" (Diamantis, 2021, p. 801). Currently, when one of their algorithms causes harm to a website user, website operators might try to argue that despite their belief otherwise, the algorithm evolved in ways that introduced defects and so this isn't the operators' responsibility (Villasenor, 2019).

To avoid this problem, Diamantis (2021) notes that corporate liability law already recognizes that corporations are "people" capable of acting injuriously through their employees, because they control and benefit from their employees' work. Given this, he argues, we should similarly also be able to use corporate liability law to attribute algorithmic harms to any corporation that (a) exercises a high degree of control over the algorithm(s) in question, and (b) lays substantial claim to the productive benefits of the algorithm(s). This is because in these cases corporations can be viewed as acting through the algorithms, which become extensions of the corporations as 'persons'.

Diamantis (2020, p. 917) goes further, arguing for an "extension of the corporate *mind* from human employees to automated algorithms". He suggests that such extension happens when (1) a "corporation knows information embedded in its algorithms"; (2) the "information is available and the employee/algorithm (on behalf of the corporation) typically invokes" it; (3) the "employee/algorithm (on behalf of the corporation) more or less automatically endorses the information upon retrieval"; (4) the "employee/algorithm (on behalf of the corporation) can easily access the information"; and (5) the algorithm uses the "information in a way that accrues some [...] benefit to the corporation" (Diamantis, 2020, p. 921-922).

The view that in algorithmic tracking website operators' mindreading might literally extend into personalization algorithms helps develop and offers additional support for Diamantis' points. For Diamantis doesn't yet consider mindreading,[14] and his conditions (1)-(5) are more susceptible to common objections to HEC such as the "coupling-constitution fallacy" or the "cognitive bloat" critique (Adams & Aizawa, 2008; Allen-Hermanson, 2013) than the conditions I mentioned above. Specifically, condition (5) that I relied on is supported by arguments from dynamical systems theory, which advocates of HEC developed precisely to address these common critiques of HEC (Palermos, 2014). Moreover, if, as the extended mindreading view suggests, the algorithms that website operators employ to identify website users' mental states are sometimes literally part of the website operators' own mindreading and their cognitive system (and person), then any harmful effects of these algorithms (in these cases) can more directly be tied to the website operators' own responsibility than otherwise. For in these cases, the website operators would then partly *themselves* (through the algorithms) cause the harm. Holding corporations liable for the things they do through their algorithms may also prompt them to ensure their algorithms operate in socially beneficial ways. A conceptualization of mindreading that helps recognize that website operators sometimes act through their algorithms may thus also encourage corporations to exercise responsible control over these systems.

In short, no matter whether the extended mindreading that we focus on is performed by website operators (individuals/corporations) or users, the extended mindreading view provides new conceptual tools to correct the current imbalance in power and control between website operators and users with respect to the algorithms at play. It provides conceptual links to be able to use already existing ethical and legal framework concerning (a) personal assault (Carter & Palermos, 2016), and (b) corporate liability (Diamantis, 2020, 2021) to guard people from potential algorithmic harm by website operators including big Internet companies (Facebook, Google, Amazon, Twitter, etc.).

8. Conclusion

Mindreading happens inside people's head. But does it *only* occur there? Here I challenged the assumption that it does by relating mindreading to HEC and the phenomenon of algorithmic tracking. I argued that if we endorse common conditions for extended cognition, then in algorithmic tracking, the mindreading of both website operators and users is sometimes not only causally coupled to but partly realized by and so extended into the computer systems involved. This is (*inter alia*) because arithmetic tracking is a property that isn't adequately ascribed to either these computer systems or the website operators or users alone but emerges

[14] Diamantis also doesn't yet note that website *users'* mindreading, too, might extend into algorithms. Nor does he consider the point above related to existing assault legislation. This point can, however, usefully be combined with his argument.

from a system that includes both. And within this broader system, we can't clearly distinguish inputs and outputs going from one to the other interactant.

While the argument I offered here for the extended mindreading view is conditional in nature, it has several interesting implications. Since it rests on the assumption of conditions for extended cognition that are widely accepted among friends of HEC, many advocates of HEC are now committed to the notion of extended mindreading. More importantly, while there might be many objections to HEC and the notion of extended mindreading, the preceding discussion suggests that this notion also offers novel ways in which we may effectively alert people to and protect ourselves against the increasing risks of computational control and online manipulation. These points provide good grounds to take the extended mindreading view seriously.

Acknowledgements

I'm grateful to two reviewers of this journal and the members of the weekly AI seminar at the CST Bonn for very helpful feedback on the paper.

References

Adams, F., & K. Aizawa. (2008). *The Bounds of Cognition*. Blackwell Publishing.

Alfano, M., & Skorburg, J. A. (2017). The embedded and extended character hypotheses. In J. Kiverstein (Ed.), *Philosophy of the social mind* (pp. 465–778). New York: Routledge.

Allen-Hermanson, S. (2013). Superdupersizing the mind: extended cognition and the persistence of cognitive bloat. *Philosophical Studies*, 164, 791–806.https://doi.org/10.1007/s11098-012-9914-7

Andrews, K. (2012). Do apes read minds? Toward a new folk psychology. Cambridge, MA: MIT Press.

Apperly, I. (2011). *Mindreaders: The Cognitive Basis of 'Theory of Mind'*. Psychology Press.

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, 124, 150–159.

Burke, R. (2007). Hybrid Web Recommender Systems. In: Brusilovsky P., Kobsa A., Nejdl W. (eds) *The Adaptive Web. Lecture Notes in Computer Science*, 4321. Springer, Berlin, Heidelberg.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*. https://doi.org/10.1177/2053951715622512

Burr, C., & Cristianini, N. (2019). Can Machines Read our Minds? *Minds & Machines*, 29, 461–494.

Carruthers, P. (2011). *The Opacity of the Mind*. New York: Oxford University Press.

Carter, J., & Palermos S. (2016). Is Having Your Computer Compromised a Personal Assault? The Ethics of Extended Cognition. *Journal of the American Philosophical Association*, 2 (4), 542-560.

Carter, J., Palermos, S., & Collin, J. (2017). Semantic inferentialism as (a form of) active externalism. *Phenomenology and the Cognitive Sciences*,16, 3, 387-402.

Clark, A. (2008). *Supersizing the mind*. Oxford: Oxford University Press.

Clark, A. (2010). Memento's revenge: The extended mind, extended. In Menary (Ed.), *The extended mind*. Cambridge, Massachusetts: MIT Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, 1, 769–771.

Cushman F. (2019). Rationalization is rational. *The Behavioral and brain sciences*, *43*, e28. https://doi.org/10.1017/S0140525X19001730

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 1(4), 568–570.

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

Diamantis, M.E. (2020). The Extended Corporate Mind: When Corporations Use AI to Break the Law. *North Carolina Law Review*, 98, 4, 6: 894-932.

Diamantis, M. (2021). Algorithms acting badly: A solution from corporate law. *George Washington Law Review*. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3545436

Drayson, Z. & Clark, A. (2020). Cognitive disability and embodied, extended minds. In the *Oxford Handbook of Philosophy and Disability*, (eds) Wasserman, D. and Cureton, A. OUP.

Fitz, N., and Reiner, P. (2016). Perspective: Time to expand the mind. *Nature* 531, S9, https://doi.org/10.1038/531S9a

Gallagher, S. (2013). The socially extended mind. *Cognitive Systems Research*, 25-26, 4–12.

Giere, R.N. (2007). Distributed Cognition without Distributed Knowing. *Social Epistemology*, 21: 3,313-320, DOI: 10.1080/02691720701674197

Gomez-Uribe, C. & Hunt, N. (2015). The Netflix recommender system: Algorithms, business value, and innovation. *ACM TMIS* 6, 4 (2015), 13, 1–19.

Halina, M. (2015). There is no special problem of mindreading in nonhuman animals. *Philosophy of Science*, 8, 2 (3), 473-490.

Hao, K. (2021). The Facebook whistleblower says its algorithms are dangerous. Here's why. *MIT Technology Review*. URL:

https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms/

Hardesty, L. (2017). Try this! Researchers devise better recommendation algorithm. *MIT News*, 1–8.

Heersmink, R. (2017). Extended mind and cognitive enhancement: moral aspects of cognitive artifacts. *Phenom Cogn Sci* 16, 17–32. https://doi.org/10.1007/s11097-015-9448-5

Heersmink, R., & Sutton, J. (2020). Cognition and the web: extended, transactive, or scaffolded? *Erkenntnis*, 85(1), 139–164.

Hernandez-Orallo, J. & Vold, K.V. (2019). AI Extenders: The Ethical and Societal Implications of Humans Cognitively Extended by AI. *Proceedings of the AAAI/ACM 2019Conference on AIES*, 507-513. URL: https://dl.acm.org/doi/pdf/10.1145/3306618.3314238

Hinds, J., & Joinson, A. (2019). Human and computer personality prediction from digital footprints. *Current Directions in Psychological Science*, 28(2), 204–211.

Lambiotte, R. & Kosinski, M. (2014). Tracking the digital footprints of personality. *Proceedings of the Institute of Electrical and Electronics Engineers*, 1935–1939.

Locke, J. (1967). *Two Treaties of Government*, P. Laslett (Ed). Cambridge: CUP.

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art [Editorial]. *British Journal of Developmental Psychology*, 30(1), 1–13.

Lynch, M. (2013). *Brief of Michael P. Lynch as Amicus Curiae in Support of the Plaintiffs, American Civil Liberties Union v. James R. Clapper.* Available from: https://www.aclu.org/legal-document/aclu-v-clapper-amicusbrief-michael-p-lynch-philosophy-professor-university

Lynch, M. (2019). Know-It-All Society: Truth and Arrogance in Political Culture. NY/London: Liveright/W.W.Norton & Company.

Malle, B. F., Moses, L. J., & Baldwin, D. A. (2001). Intentions and intentionality. Foundations of social cognition, Cambridge, MA: MIT Press.

Nathani, D., Chwastek, M. & Sreenivas, V. (2020). Amazon Personalize improvements reduce model training time by up to 40% and latency for generating recommendations by up to 30%. *Amazon Web Service Machine Learning Blog.* URL: https://aws.amazon.com/blogs/machine-learning/amazon-personalize-improvements-reduce-model-training-time-by-up-to-40-and-latency-for-generating-recommendations-by-up-to-30/

Nichols, S., & Stich, S. P. (2003). *Mindreading*. Oxford: Oxford University Press.

Oremus, W. (2016). Who Controls Your Facebook Feed. *Slate.* URL: http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html?via=gdpr-consent&via=gdpr-consent

Palermos, S. O. (2014). Loops, Constitution, and Cognitive Extension. *Cognitive Systems Research* 27: 25–41.

Palermos, S. O. (2016). The dynamics of group cognition. *Minds and Machines*, 26(4), 409–440.

Palermos, S., & Tollefsen, D. P. (2018). Group Know-How. In J. A. Carter, A. Clark, J. Kallestrup, S. O. Palermos, & D. Pritchard (Eds.), *Socially Extended Epistemology*, Oxford University Press.

Peters, U. (2019). The Complementarity of Mindshaping and Mindreading. *Phenomenology and the Cognitive Sciences* 18 (3): 533-549.
Peters, U. (2021). Objectivity, Perceptual Constancy, and Teleology in Young Children. *Mind &Language*. https://doi.org/10.1111/mila.12344

Phillips, J., & Norby, A. (2019). Factive theory of mind. *Mind & Language*. https://doi.org/10.1111/mila.12267.

Rabinowitz, N., Perbet, F., Song, F., Zhang, C., Eslami, S.M.A. & Botvinick, M.. (2018). Machine Theory of Mind. *Proceedings of the 35th International Conference on Machine Learning*, 80, 4218–4227.

Reiner, P. B. and Nagel, S. (2017). Technologies of the Extended Mind: Defining the Issues. In *Neuroethics: Anticipating the Future*, J. Illes, ed. pp. 108-122 (2017), Available at SSRN: https://ssrn.com/abstract=2961342

Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (2011) (eds). *Recommender Systems Handbook*, Springer, Berlin.

Rupert, R.D. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy*, 101(8), 389–428.

Rust, J., Kosinski, M, & Stillwell, D. (2021). *Modern Psychometrics. The Science of Psychological Assessment*. Routledge.

Schwengerer, L. (2021) Online Intellectual Virtues and the Extended Mind. *Social Epistemology*, 35: 3, 312-322, DOI: 10.1080/02691728.2020.1815095

Smith, B. & Linden, G. (2017). Two decades of recommender systems at Amazon. *Com, IEEE Internet Computing*, 21, 3, 12–18.

Spaulding, S. (2020). What is mindreading? *WIREs Cognitive Science*, 11: e1523. https://doi.org/10.1002/wcs.1523

Sprevak, M. (2020). Extended cognition. In: Crane, T. (Ed.) (2019). *The Routledge Encyclopedia of Philosophy Online*, London, Routledge. Retrieved from: https://marksprevak.com/pdf/paper/Sprevak--- REP%20Extended%20Cognition.pdf

Theiner, G., Allen, C., & Goldstone, R. L. (2010). Recognizing group cognition. *Cognitive Systems Research*, 11(4), 378–395. doi:10.1016/j.cogsys.2010.07.002

Thornton, M. A., & Mitchell, J. P. (2017). Theories of person perception predict patterns of neural activity during mentalizing. *Cerebral Cortex*, 28(10), 3505–3520.

Villasenor, J. (2019). Products liability law as a way to address AI harms. *Brookings Report*. URL: https://www.brookings.edu/research/products-liability-law-as-a-way-to-address-ai-harms/#footnote-3

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences of the United States of America*, 112(4), 1036–1040.

Zanker, M., Rook, L., & Jannach, D. (2019). Measuring the impact of online personalisation: Past, present and future. *International Journal of Human-Computer Studies*, 131, 160–168.

Zhang, Y., Feng, F., Wang, C., He, X., Wang, M., Li, Y., & Zhang, Y. (2020). How to Retrain Recommender System?: A Sequential Meta-Learning Method. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. https://arxiv.org/abs/2005.13258