Synthese (2025) 205:133
https://doi.org/10.1007/s11229-025-04962-3

ORIGINAL RESEARCH

The extracted mind

Louis Loock[1]

Received: 12 July 2024 / Accepted: 11 February 2025

Abstract

Since Clark and Chalmers advanced "The Extended Mind" in 1998, a persistent dispute evolved on how our tool interactions shape the kind of cognition we have. Extended cognition generally views us as cognitively augmented and enhanced by our tool practices, which shall render our cognitive constitution extended to those tools. Bounded and embedded cognition have primarily criticized this metaphysical claim. However, another contender may arise from considering how we use more intelligent tools. We arguably employ advanced technologies that capture, mimic, and then replace our cognitive skills, which we then no longer need to exercise ourselves. This precedes any metaphysical debate, since such practices might stand in a more fundamental conflict with extended cognition. The counter-hypothesis of extracted cognition states that we primarily tend to use tools that initially attain and eventually displace our cognitive responsibilities and involvements. This paper evaluates extended and extracted cognition by comparing theoretical, practical, and ethical arguments respectively. If extracted cognition describes most convincingly how such tool interactions shape our kind of cognition, then we may also endorse "The Extracted Mind".

Keywords Situated cognition · Extended cognition · Cognitive tools · Tool practices · Cognitive impact · Artificial intelligence

✉ Louis Loock
louis@loock.pro

[1] Institute of Cognitive Science, Osnabrück University, Wachsbleiche 27, 49090 Osnabrück, Germany

Published online: 13 March 2025

1 The debate

What kind of cognition do we have? Each philosophy of cognition advances a different conceptual answer to this plain question. Situated cognition research and its philosophy (Newen et al., 2018; Robbins & Aydede, 2008) address a specialized question: What kind of cognition do we have, given how *we* situate our cognizing with environmental factors? This addition shows how situated theorizing conceptualizes different kinds of cognition relative to the unique external relations we initiate, control, and develop for our own cognitive goals. Situated answers envision the very nature of our cognition as, say, embodied, embedded, extended, enacted or (socially) distributed—any metaphysical claim about the location of our cognitive vehicles ensues from that. The present paper is therefore not only interested in cognitively profound organism–environment interactions, but primarily how they shape the kind of cognition we ultimately have.

Especially the philosophical debate stemming from Clark and Chalmers' seminal push for extended cognition (1998) has prominently depicted our cognizing as a self-situating activity of internal processes over external artifacts (cf. Slaby, 2016, who calls this the "user/resource model"). This tendency of thought views cognitive agents as causally controlling and epistemically exploiting external structures for their own cognitive benefits. Vehicle externalists (e.g., Clark, 2008; Hurley, 1998; Menary, 2007; Rowlands, 2010; Shapiro, 2019; Wheeler, 2005; Wilson, 2004) are claiming that, because of this interaction style, cognition is constitutively extended over intra- and extra-organismic vehicles.

Virtually all theorizing of early cognitive science bequeathed this 'agent-in-charge' understanding to newer situated cognition research and its critique to follow. So far, neither proponents nor contesters of extended cognition faced any reason to unravel and challenge this understanding. Adams and Aizawa defend the bounds of cognition (2001, 2008a, 2010a) to argue that any external interactions are mere causal accompaniments of our cognition which, as a natural kind, only comprises those more homogeneous neural processes within the organism. Rupert explores the middle ground of embedded cognition (2004, 2009) to emphasize how our inner cognitive processes are highly adaptive to, but not co-constituted by, external structures and tools that we use to solve our cognitive tasks more easily.

In all discussed cases of this debate, the agent is unquestionably viewed as the most capable and active component of any extended, embedded, or bounded cognitive system. There just seem to be no tools that steer the cognitive interaction more dominantly than cognitive agents can and in fact do. However, this presupposition may be challenged by considering recent and expectable advancements of artificial intelligence (AI) and its implementation in our digital tools. Another kind of situated cognition may form when our cognizing is less shaped by our directing of external tools, but by the way these highly capable gadgets redefine the cognitive tactics we consider or discard from the start. We may employ new technologies not to assist and augment, but rather to absorb and displace our cognition. This may happen when we forward our cognitive skills to these much more powerful tools that can solve our cognitive tasks much quicker, better, and cheaper for us. Therefore, a new critique of extended cognition may arise by letting these new interactions and their cogni-

tive impacts lead us to a strongly opposing conclusion. The counter-hypothesis of extracted cognition claims that cognitive agents seek, and may actively transfer their strategies to, cognitive artifacts such that these take over a considerable burden of processing, and so the cognitive involvement of the agent becomes redundant and dispensable for the given task. If our cognitive practices actually develop this way, then extended, embedded or bounded cognition all do not capture this novel situated phenomenon, and so a new type of situated cognition is needed.

To develop such a seemingly bizarre hypothesis, we shall proceed as follows. Section 1 paved the way for situated theorizing to commence with an alternative perspective: What kind of cognition do we have, given how *environmental factors* situate our cognizing? Section 2 briefly reviews a classic example of extended cognition to unravel important externalistic assumptions that shall be critically reviewed. Section 3 then considers two different cases to reflect on these assumptions accordingly, and to thereby motivate the transition from extended to extracted cognition. Section 4 has three subsections that analyze and compare the theoretical, practical, and ethical arguments for extended and extracted cognition respectively. Section 5 responds to some potential critique in order to establish this new hypothesis as an eligible contender.

2 The assumptions

The extended cognition literature has proposed many cases in its alleged favor. However, reapproaching just one paradigmatic scenario suffices to get an intuitive first grasp of its hypothesis. In the following, the example of Otto is used to explain a fundamental principle (of situated theorizing in general) and three disputable assumptions (of extended cognition in particular). Section three then discusses two other examples that turn these assumptions around and provoke intuitions for extracted cognition instead.

Otto, an Alzheimer's patient, uses his notebook as an external, compensatory memory structure (Clark & Chalmers, 1998, pp. 12–18). While we normally conceive of memorizing as exclusively realized by specific neural processes inside of us, interacting with a special external tool in our cognitive routine might satisfy the same cognitive function as some internal processes usually would. Memory encoding might be realized by transferring information into a suitable external structure; memory storage is established by the information rigidity of the structure and by the agent's protection and cleanup strategies for it; and lastly, memory retrieval occurs whenever the agent is recollecting the relevant pieces of information from that external structure. Overall, this tool use strategy greatly improves his cognitive capabilities and autonomy. By externalistic standards, the notebook then constitutes an extra-organismic cognitive vehicle of Otto's memory. This kind of memorizing is one of his extended cognitive processes and so the notebook becomes a part of his extended cognitive system.

Independent of the medical condition, many people may be considered extended cognizers, too. For example, whenever we use our phones to remember things in a similar fashion: we type, forget, and get reminded. Yet, Otto's pathological state was purposefully used by Clark and Chalmers to pump moral intuitions for calling Otto

an extended cognizer (see Sect. 3). Though to not let this aspect pollute independent argument types for this hypothesis (see Sects. 4.1 and 4.2), we shall also imagine Otto as neurologically healthy to evaluate how this impacts our cognitive judgment.

Now, what is usually driving us to seek and establish such environmental dependencies for our thinking? Our cognition comprises causal processes that are generally resource costly, yet the task constraints we face are demanding resource efficiency, especially in their totality. Luckily, we often have several options for approaching our cognitive tasks—some purely internal and others partly, or fully, external. Each option comes with a different profile of costs and benefits for us. Extended cognition conceived a basic principle for such decisions: Take the most lucrative resource, no matter its location, as long as it brings you the desired outcome. Clark coined this first the 007-principle (1989, p. 64) and later the hypothesis of cognitive impartiality (2007, p. 174), whereas Rowlands calls it the barking dog principle (2010, pp. 15–16). This principle may however be expressed more accessibly by the notion of a *cognitive economy* (cf. Clark, 2007, p. 185): What is available at what price, and what can and should I afford? Otto's cognitive employment of a notebook is economically reasonable in this sense, because he cannot spend much internal resources for such tasks himself, and the external storage of information is usually deemed more reliable, robust, and simpler. More generally then, our choice of cognitive tactics is fundamentally driven by optimizing the cognitive business we manage for all available resources. Now, extended cognition uses this principle to emphasize how the cognitive resources we actually purchase lie outside our organismic system more often than cognitive scientists and philosophers commonly assumed.

Beyond this basic principle, the following three assumptions are more crucial for extended cognition in particular. However, we shall not view these assumptions as decisive criteria for that position (see Sect. 4 for that), but rather as pre-theoretical and intuitive descriptions of what impels extended theorizing. For this purpose, these assumptions further utilize the metaphor of a cognitive economy.

(1) *Cognitive worker*. Extended cognition generally retains the well-established assumption that special intra-organismic structures (perhaps only of our nervous system) are primarily steering our cognitive activities. Even when external props extend the scope of our cognitive system at the periphery, it is still these central structures that are causally controlling and epistemically exploiting the external tools to decide about the cognitive task in focus. As Clark repeatedly phrases it in his updated hypothesis, our cognition is not organism-bound, but still organism-centered (Clark, 2007, p. 192, 2008, p. 139, 2010c, p. 1059; cf. critique by Menary, 2010c). So while some cognitive processes, skills, and outcomes are only due to, and only manifest with, the usage of specific tools (Rowlands, 2009, pp. 3–4), the internal processes still carry most of the processing burden for starting, developing, and completing any cognitive solution (Rowlands, 2010, Chap. 1). We may therefore say that Otto, not his notebook, is the prime cognitive worker in this interaction, as he decides what to write down when and where, what to look up again, and how to use that information to come up with further ideas.
(2) *Cognitive artifact*. While it is generally recognized that, so far, no external tools or structures are cognitive on their own (Adams & Aizawa, 2010a; Clark, 2010a;

Menary, 2006, pp. 332–333; Rowlands, 2010, pp. 92, 127), externalists argue that our tools can become cognitive components of a larger cognitive process and system whenever we, as cognitive agents, procedurally couple to, or integrate them into our cognitive routines for momentary cognitive demands (cf. Clark, 2007, pp. 183–185). Still, the tools themselves are, if at all, very lean in cognitive regards and users need to transform their own skills so to then appropriately manipulate the tool for any cognitive achievements to follow. In this way, Otto's notebook can at most be considered a cognitive artifact when appropriately used (notion introduced by Norman, 1991; also see Heersmink, 2013; Fasoli, 2018), because it does not start and maintain any cognitive procedures itself, yet can be honed and exploited for such by the agent. These first two assumptions fix a special imbalance present in all discussed cases of extended cognition: the agent carries most of the processing share and responsibility, while the tool itself is a rather passive, purely mechanical, and non-adaptive artifact.

(3) *Cognitive profits*. Lastly, the impact of our tool interactions can be evaluated from two views: a personal view and a systemic view (cf. Norman, 1991). Extended cognition generally depicts our external cognitive interactions as creating both personal and systemic cognitive benefits for us. So allegedly, using any tool in a cognitively productive way (systemic profit) should usually also improve our corresponding tool-use skills (personal profit). When the external cognitive economy flourishes, then our inner cognition writes profits as well. This assumption of extended theorizing has only recently gained explicit attention (Aagaard, 2020; Walsh, 2017). This is somewhat surprising given that "advanced cognition" initially labeled a pivotal motivation for the consequent debate (Clark, 1997, p. 180). The example of Otto is a remarkable model for this idea, since the constant use of his notebook both greatly amplifies Otto's overall memory capacities (systemic profit), but it also enhances many of his individual skills needed for an effective and efficient interaction (personal profit). If this conception is now applied to all tool-using individuals, then we understand why some externalists tell the human history of tool developments really as the story of cognitive advancements (cf. Clark, 2003). Yet, a normative stance of tool-use optimism, deriving from these personal and systemic impacts, should be kept in conceptual separation from this assumption.

Conjoining these assumptions with the principle of a cognitive economy brings the extended vision of our cognition into full view. We find ourselves in a world full of cognitive tasks and many options to approach them. While we remain in charge of the cognitive interaction in general, we are in the habit of exploiting external resources that assist and augment our thinking in the details. Ultimately, our tools make us think more and better in any case—at least in the extended framework.

3 The problem

Are there scenarios where these externalistic assumptions misinterpret the nature of our tool-dependent cognizing? If extended cognition wants to defend its integrity and credence, all three assumptions should be equally applicable to novel cases. Otherwise, we may need to revise those assumptions and conceive a more suitable kind of cognition for such circumstances. The present section assesses these assumptions with two more advanced tools and their employment in our cognitive economy: pocket calculators and AI-powered devices.

Pocket calculators might just as well be candidates for cognitive extension. When used appropriately, the cognitive strategy for solving a mathematical task, as a whole, might be constituted by processes which are spread across internal and external vehicles—at least according to some externalistic frameworks (cf. Clark & Chalmers, 1998, p. 11). What certainly remains fixed for this example, too, is the principle of a cognitive economy: rerouting a central step of our mathematical solution approach into a calculator is—for most people and in most settings—more lucrative, since it makes the calculation much faster, simpler, and more reliable. This principle therefore marks an uncontested cornerstone, and only beyond it does any disagreement unfold. So before deciding for one hypothesis over the other, all discussed cases entail some kind of *cognitive reallocation* between internal and external structures: some steps in our inner processing chain (biological retention and calculation) are rendered obsolete when they are taken over by external surrogate processes (in the notebook or calculator). However, revisiting the three externalistic assumptions reveals important differences.

Deviation from (1). Already the use of a calculator may bring some doubt about our status as cognitive workers. While the calculator cannot approach and solve our math problems on its own, we can make it take over the calculation part so that we have to think about one subprocess less in the overall cognitive procedure. From this systemic perspective, the calculator relieves us from a more complex responsibility than a notebook usually does. The cognitive burdens across organism and tool change more drastically in the calculator case, as we are suddenly no longer responsible for the most crucial and most demanding step on the solution path. Heersmink (2012, p. 53) analyzes this dimension of cognitive interactions as the "distribution of computation" and Fasoli (2018) consequently describes such tools as "substitutive cognitive artifacts", which others adopted (e.g., Cassinadri, 2024, p. 11). The second deviation covers the other side of this change in balance.

Deviation from (2). Simply typing an equation into a calculator will activate internal mechanisms that complete the calculation for us. To repeat, those mechanisms are much more sophisticated than anything a plain notebook could offer, because the outputs it returns are much more demanding to achieve than simply leaving our inputs untouched, as standard notebooks would. Therefore, we may feel a slight sense of unease when considering our calculator *in use* a mere cognitive artifact (as specified above), since it carries a much more demanding cognitive responsibility for us. However, to already consider the calculator as a cognitive worker in such interactions may also be too charitable. It is at least fair to say that the cognitive balance is shifting,

perhaps to some kind of cognitive equilibrium—though externalists would not yet take this as speaking against their proposal (cf. Aagaard, 2020, p. 175).

Deviation from (3). Finally, there is some persisting controversy about the cognitive impacts (profits versus costs) of such tools. From the systemic perspective, admittedly, the cognitive deal in fact generates cognitive gains far beyond what individualistic processes can usually offer on their own (measured in the speed, accuracy, reliability, and quantity of solved equations). But from the individual perspective, forwarding parts of a cognitive task to a more capable tool decreases both the actual work done, and also excludes some learning potential that would usually be imposed by solving such tasks without external assistance. We learn using the tool (faster), but not the primary skill otherwise required. A strict cognitive economy can certainly increase the overall cognitive profits, but they usually come with the cost of an increased individual dependency on the tool in turn, especially in the long run. The more we think with our tools, the less we desire to think without them. Using a calculator will not immediately deprive us of the skill of doing calculations ourselves—that is clear—, but it also cannot possibly boost our individual fluency and capacity for those tasks. This controversy about the cognitive impacts of calculator usage is sufficient to point us at the present deviation.

While this might have triggered some unease, a new antagonist is needed to provoke a full change of mind. Imagine Anna, the daughter of Otto, who works as a digital marketing director at the Museum of Modern Art (MoMA). We will not portray her with any medical condition for now, because (as said for the Otto case) this dimension shall only come in where it actually matters (see Sect. 3). For her completely digital and remote position, she was instructed to use a laptop with a state-of-the-art operating system and all kinds of applications that are powered by an AI. Although such AI-powered devices are ultimately still realized by rather brute and unreflective mechanisms as we find them in any calculator, there is at least a significant quantitative difference. Most current AI applications primarily perform enormous statistical procedures on tons of training and verification data so as to eventually yield impressive generative results in the desired modalities. Mühlhoff explicated an inspiring understanding of what must happen on a societal scale to enable these widely admired algorithmic achievements: "[Machine Learning] is more than algorithms and [High Performance Computing]: it is a media-cultural constellation involving human-machine interfaces and media technology that makes people implicitly generate data that can be used as training data." (Mühlhoff, 2020, p. 1874). If we then look at the individual level again, we might conceive how Anna's laptop is built and employed to (additionally) source some of its artificial intelligence directly from her natural intelligence. More technically, the parameters of these models would be fine-tuned toward the provided personal data of Anna to then produce outputs of similar style and content. How to understand this cogno-technological innovation?

Suppose Anna's laptop has a unified user interface that only presents itself via a digital assistant which can access and direct all other programs of the underlying system (cf. Clark, 2014, pp. 186–187). This is not science fiction, but consumer reality of tomorrow—at least if we slightly extrapolate the current developments of AI and its implementation in our existing digital devices. Now imagine the moment when Anna boots her laptop for the very first time. Many personal questions come up about

her job, duties, strategies, workflow, attitude and what not. Given Anna's data and instructions, her new digital assistant becomes able to learn about and take over many of the digital tasks Anna would usually solve with her individual cognition: mail writing, calendar updating, graphic design, coding, researching, video and photo editing, reporting about new projects, translating messages, presenting her face and voice in digital conversations, and it might even learn to reason like her—seemingly at least. Eventually, Anna can simply prompt all her ideas and duties to this one assistant and it will automatically select the required subroutines for the task at hand. The market of cognitive goods will have a new vendor, even though it need not (and does not) deploy cognitive means of production.

To first reconfirm this, Anna sticks to the principle of a cognitive economy, since her laptop usage frees up lots of time and energy whenever she just bluntly prompts all her questions, ideas, wishes, notes, or tasks firstly and directly to the AI application. This interaction trend is strengthened further when both Anna is eager to work and think less, and her tool is equally incentivizing her for that tactic and so offers to take over all possible tasks to the point of local automations. This is what the AI tool is designed to do: boosting the quality and quantity of the working outputs, while decreasing Anna's cognitive involvement and responsibilities as much as possible. This scenario can subvert the three externalistic assumptions and thereby motivates the alternative framework of extracted cognition.

Subversion of (1). Since Anna started working in this way, she is significantly less involved with many aspects of her work, because the AI assistant is now dealing with them (cf. Hernández-Orallo & Vold, 2019). If well-disposed, we could say that she promoted herself from a cognitive *worker* to a cognitive *manager* of all these AI-generated results. Her AI can still receive some guiding feedback and Anna might loosely check its final outputs, but she no longer has to attend the cognitive demands of every task (cf. Cassinadri, 2024, pp. 11–12).

Subversion of (2). In turn, her AI assistant and its subprocesses are actually producing most of Anna's working outcomes—the posters, code, mails, texts, images or videos. It may even download further programs to conveniently relieve Anna from further tasks. It also analyzes and adapts to Anna's inputs so that, eventually, the AI is not only wielding more decisional power in that interaction, it also seems to imitate many cognitive habits of Anna. While the laptop and its processing units still perform billions of perfectly rigid, unreflective, and non-cognitive transistor switches, this package may nevertheless deserve, in this interaction, a greater procedural or even *epistemic credit* for its contributions to so many of Anna's tasks (cf. Clark & Chalmers, 1998, p. 8). We can therefore no longer consider the laptop a mere cognitive *artifact*, since it got promoted to the cognitive *worker* of this interaction.

Subversion of (3). The externalists might still point at the systemic cognitive gains through this interaction (i.e., more tasks are completed with less effort and faster), but Anna's individual cognitive costs would be blatantly overlooked and traduced. Only considering the system view arguably misses the actual question in focus: What kind of cognition should we ascribe *to Anna* in light of her tool use practice? Since extended cognition claims both systemic and individual cognitive improvements, it does not apply to Anna's situation. While her practices are very profitable for the overall working output, their excess may come with considerable costs for Anna's

individual cognition. So there is at least a cognitive tradeoff that extended cognition does not properly address (Cassinadri, 2024, p. 14; Fasoli, 2018, pp. 684–685), and more destructive consequences seem not unlikely for the long run (see Sect. 4.2). But as said above, any normative evaluation is not implied by this descriptive point.

Other assumptions could have been argued with, but these three in particular already suffice to introduce a novel perspective. By always pursuing a highly efficient cognitive economy, we now see why these externalistic assumptions need to be updated with regard to the usage of advanced tools or other structures: we might become the distant managers of cognitive tasks that our working tools solve *for us*, but largely *without us*, and so our cognitive condition can no longer be viewed as beneficially 'extended', but should be seen as deleteriously 'extracted'. While the hypothesis of extended cognition states that external processes may at times also function as compensating surrogates in our extended cognitive systems, they only ever do so in compatibility and without competition to our working internal processes. The presented alternative suggests that AI applications are in fact no such cognitive extenders, but are rather employed as cognitive extractors—even if they do not preserve our actual cognition, they may still function as digital emulators of it. The hypothesis of extracted cognition envisions a cognitive reality in which we captivate ourselves in the practice of letting all our smart devices first capture, then mimic, and eventually replace the cognitive skills we previously only executed over our own internal processes. Calling these adversaries 'hypotheses' is appropriate and follows the canon of the previous debate, since a full theory of cognition is thereby not stated nor strictly imposed, though they go beyond merely pointing at possible cognitive interaction strategies by construing a vision for the kind of cognition we thereby possess. To summarize their difference: While extended cognizers seek external tools that still require the co-activation of relevant internal processes to solve a task, extracted cognizers are surrounded by tools that can, should, and do eventually solve the task without any of our internal contribution.

4 The arguments

Now that the ideas of both extended and extracted cognition are intuitively introduced, we can (re-)construct more concrete arguments for either hypothesis, that is, theoretical, practical, and ethical arguments in particular. In this order, the following three subsections explicate *constitutive criteria* (What constituents and causes define the kind of cognition we have?), *teleological criteria* (What goals and consequences define the kind of cognition we have?), and *axiological criteria* (What values and normative expectations define the kind of cognition we have?). Each of these sets of criteria may be combined to support their corresponding hypothesis, but here we focus on the conflicts between extended and extracted explications of every argument type individually. Hence, each upcoming subsection first reconstructs the respective argument type for extended cognition in its strongest conceivable form, before additional or countering criteria for extracted cognition are introduced. By this tactic, we pay a fair tribute to the extended cognition literature, and still show how countering or additional aspects carry us to an opposing cognitive conception. As before, Otto is

posed as the extended and Anna as our extracted archetype. Note, however, that some variations applied to these examples may override the kind of cognition we would intuitively ascribe based on the presented criteria (e.g., their medical condition).

4.1 Theoretical arguments

The breadth of theoretical considerations for extended cognition has become too diverse to be exhaustively systematized in just one paper. Externalists pioneered conceptual (e.g., Menary, 2007), methodological (e.g., Rowlands, 2010), explanatory (e.g., Clark, 2008), metaphysical (e.g., Rockwell, 2005), and also phenomenological (e.g., Wheeler, 2005) answers to it (not to mention cross-attributions). The following, however, only covers metaphysical criteria, as these are the most widely discussed.

Critics have divided the metaphysical thrusts for extended cognition in the following way: Some externalists allegedly only give *causal criteria* without explicitly proposing *constitutive criteria* of cognition. This verdict stems from Adams and Aizawa's ubiquitous critique of a 'coupling–constitution fallacy' and their call for a 'mark of the cognitive' (Adams & Aizawa, 2001, 2008a, b, 2010a, b; Adams, 2010; Aizawa, 2010). Consequently, some externalists (Rowlands, 2009; cf. Menary, 2006, pp. 330, 334) compliantly and explicitly address the constitutive *what*- before the causal *where*-question of cognition (cf. Walter & Kästner, 2012).

However, this crude separation between causal and constitutive criteria of (extended) cognition is rather misleading. Although the search for a mark of the cognitive is strictly required to conclusively settle the question of metaphysical constitution, it must be supplemented by relevant causal criteria about the interaction between organism and environment (at least within the situated framework). The reason for this is that 'cognition' shall delineate a scientific (though not necessarily natural) kind of causal processes in the world (Menary, 2010c, p. 608). Therefore, the relevant causal facts need to be identified and then appropriately abstracted to form causal criteria for the cognitive kind (Clark, 2010a; Piredda, 2017). This does not need to fix the cognitive analysis and study on a very low causal level (cf. Clark, 2010b, pp. 50–51), but it should make us reapproach the interaction-inspired approaches that some externalists initially developed (Clark & Chalmers, 1998 being the paramount example) without putting too much emphasis on highly theoretical conditions beforehand.

Therefore, the fittest criteria of the extended cognition literature, be they of the causal or the constitutive kind, shall be combined to construe the best possible version of extended cognition. We thereby avoid randomly choosing or deliberately construing an extended account that might be, it could be criticized, not representative enough or too weak a target to begin with. In the following, all collected criteria are grouped into three distinct stages, each of which establishes increasingly stronger sub-theses. Together, the stages shall ultimately form an abductive, that is, defeasible inference to the hypothesis of extended cognition. But given that the first three stages do not yet induce any pertinent divide between the Otto and the Anna case, a fourth stage is needed to evoke a genuine competition between those hypotheses. This last stage consists of three novel criteria that explicate why Otto is an extended and Anna an extracted cognizer.

Stage one: coupled system. We initially need to establish that there are profound causal relations between internal and external processes which justify conceiving a causally *coupled*, *integrated*, or *widened* system. First, there needs to be a rather strong bidirectional causal interaction between internal and external entities. Clark and Chalmers submit multiple expressions for that, such as coupling, causal synchronicity, inside-outside continuity, or a causal loop. Clark later specified this as the deliberate recycling of bodily outputs as new perceptual inputs (e.g., 2007, p. 185). Alternatively, Menary expresses the idea in terms of causal integration (2006, 2007), which is supposed to be more restrictive, although perhaps less clearly defined. For the sake of brevity, we may subsume these different expressions under the same notion of strong causal interaction. Second, all the components of this system—organism plus tools—play an active causal role (Clark & Chalmers, 1998, p. 8; Menary, 2010b, p. 3). They are not just passive bystanders of the main causal workings, but their active involvement is invaluable for it. This is usually spelled out in terms of counterfactual causal relevance (as a minimal requirement): changing or even eliminating any important component in the system would imply severe effects for the coupled system and its behavior. Third, this strong causal relatedness needs to be fairly present in most situations and reliably protected from disconnection (Clark & Chalmers, 1998, p. 11). These three causal criteria shall indicate a causally widened system of some kind—but of which kind exactly?

Stage two: cognitive system. Now we need to argue that this coupled whole constitutes a cognitive system. First, there is eventually no way around a mark of the cognitive for this philosophical debate (Adams, 2010; Adams & Aizawa, 2001, 2008a; Adams & Garrison, 2013; Varga, 2018; Walter, 2010), though we need to leave this criterion as a placeholder here. In any case, such a mark shall be consistent with the other conditions presently under discussion. Second, the so-called *portability* or *Naked Mind* criterion demands that cognitive skills are diversely applicable to many tasks in very different circumstances. Clark and Chalmers (1998, pp. 10–12) use this to argue that any external structure or tool becomes a "core cognitive process" as soon and as long as it is employed in such a general cognitive way. Third, one version of extended cognition is famously motivated by the parity principle (Clark & Chalmers, 1998, p. 8; cf. Clark, 2007, p. 167, 2010c, p. 1050). After many discussions of this idea (e.g., Menary, 2006; Walter, 2010; Wheeler, 2011), we can express its key motivation as follows: If a process is conventionally or reasonably deemed cognitive, then this process holds on to its cognitive status irrespective of its location (cf. Sprevak, 2009, Sect. 1). In short, no process has any cognitive primacy in virtue of its location alone and the cognitive status shall be decided independently of the location. What matters for comparing an established with a candidate process is functional parity or equivalence. Therefore, this branch of extended theorizing is adhering to some kind of coarse-grain functionalism to buttress cognitive extension (Shapiro, 2008; Sprevak, 2009; Wheeler, 2010; Drayson, 2010; Walter, 2010). Fourth, another kind of extended cognition rather investigates the complementarity between agent and tools, so the way they form a productive cognitive fit despite their dissimilarities in makeup (e.g., Rowlands, 1999; Menary, 2006; Sutton, 2010). The conflict between those last two, seemingly opposing, criteria is however best resolved by uniting them (Rowlands, 2009, pp. 3–5, 2010, pp. 86–90; Wilson & Clark, 2008, pp.

70, 72). Complementarity, on one level of description, enables very different entities to become cognitive components of the same cognitive system (addressing the problem of *cognitive drain*: too few external things are potentially cognitive). A restrictive type of functional parity might ensure, on another level, that not just any entity can be constitutively included into the same cognitive system (avoiding the problem of *cognitive bloat*: too many external things are potentially cognitive).

Stage three: centralized system. Next up, it is often argued that extendedness is not just—relative to orthodox conceptions—proposing broadened, distributed, or widened cognitive systems, but that cognition still retains a causal-constitutive center at which all our cognitive activities start and terminate, often called *the core of cognition* (Clark & Chalmers, 1998, pp. 11–12; most prominent in Clark, 2007, pp. 190, 192, 2008, pp. 107–108; resumed by Menary, 2010b, pp. 7–8; also discussed in Rowlands et al., 2020, Sect. 5.3). We touched upon this idea already in the context of the cognitive worker assumption, but further clarification is needed. Clark and Chalmers originally used three criteria to argue for *extended beliefs* (1998, p. 17), as parts of an *extended mind*, which are commonly used to support the conceptually weaker claim of extended (and centralized) cognition as well. This gradation comes down to the scope of inquiry: 'cognition' pertains to all states and processes that transform our sensations to actions, whereas 'mind' only addresses those higher states and processes like experiences, beliefs, desires, emotions, and other consciously accessible states (cf. Clark & Chalmers, 1998, p. 12; Drayson, 2010, pp. 374–377). For the sake of simplicity, we stick to the broader and less pretentious notion of cognition, but we appreciate that claims on mind and mentality can be equally defended with more specialized examples. The following criteria are here primarily understood to be strongly indicative of this theoretical tendency toward a core of cognition, given that they all require a cognitive agent that can form an epistemic access to its external tools different from the causal connection specified in stage one. In other words: The epistemic criteria stem from a *personal,* and the causal criteria from a *subpersonal* level or perspective on cognition (cf. Rowlands, 2010). First, the agent constantly relies, in an epistemic sense, on its tools in various situations, and so their usage has a significant epistemic impact on the user. Therefore, the tools should always be available for such epistemic transfer, so much so that this criterion can be spelled out in terms of counterfactual epistemic relevance. Second, the tool and its relevant features are always directly accessible when the agent needs to epistemically access them. Third, the provided information is directly endorsed, or the suggested action immediately performed, based on trust and habit. These three criteria all require the causal and epistemic binding of the tool to the agent, and so the agent qua organism is the core of this widened cognitive system. Critique of this centralization in extended cognition, i.e., the implicit or explicit assumption of a pre-formed cognitive agent (cf. Rowlands, 2010, Chap. 6) comes from Menary (2006, p. 333) and his take on extended cognition that he calls *integrationism* (Menary, 2007). Even if this version turns out to be genuinely different from other expositions of the externalistic idea, the present critique could be easily specialized to that version as well (perhaps by calling it *disintegrationism*: the agent–tool interaction has the consequence of disintegrating certain cognitive skills away from the agent toward the tool). Yet, a broader

aim is envisioned here by proposing and developing extracted cognition as a coequal hypothesis against all versions of vehicle externalism coalesced in these three stages.

Stage four: extended or extracted system. There is a final criterion of the extended cognition literature that can now, when developed further, adjudicate between Otto's and Anna's case. Initially, Clark and Chalmers (1998, p. 17) also proposed the, later much criticized, past-conscious-endorsement criterion. But that one put too much emphasis on consciousness and is no longer endorsed by Clark (cf. 2010b, p. 46). A more general alternative was discussed in Rupert's critique of it (2004, pp. 402–403) in terms of personal tailoring, or by Rowlands (2010, Chap. 6) in terms of cognitive ownership. So there is some consensus that the connection between agent and tool need not only be clarified in causal, cognitive, and epistemic regards (as in the previous three stages), but also personally: When does some external entity really belong to *my* cognitive system? Hence, what renders a (centralized) cognitive system extended is its personal integration; and what renders it extracted is its personal disintegration. We can take the intuitive considerations from above (see Sect. 3) to specify this rough notion of, say, ownership with an action-theoretical, a causal, and an epistemic criterion (to be discussed in this order). These criteria reveal unavoidably different results for Otto and Anna.

First, how much *decisional power* does the agent exert over the tool in their interaction? While Otto has to create new entries, change or erase them, and guide his consequent thoughts and actions by them, the notebook itself is not able to make such decisions for him. In contrast, Anna's AI can direct large parts of what she cognitively pursues even without her active instruction for each decision by the tool. So Otto thinks (more) through his tool, but Anna stops some of her thinking when her AI takes over (cf. Hernández-Orallo & Vold, 2019, p. 509). We could also phrase this in terms of cognitive authorship (the AI handles the largest cognitive share of some results), or even cognitive hegemony (the AI is in charge over Anna's cognitive decisions).

Second, how much *causal control* does the agent possess over the tool in their interaction? Anna only interacts in a dialogical way with the final results of her laptop, and she is not directly manipulating its underlying production processes. Otto, however, has to curate all cognitively relevant processes in his notebook. So while Otto can directly interact with the cognitive means of his artifact, Anna can only interact with the final outputs of her AI.

Third, how much *epistemic access* does the agent have to the workings of the tool in their interaction? Anna's laptop and its AI applications are an epistemically intransparent, opaque tool she has usually no momentary access to or awareness of (for a discussion of different notions of transparency in this context see Andrada et al., 2023). We might say that the epistemic ecology between tool and agent is heavily restricted or even impaired. She has not meticulously designed her AI program like Otto did his notebook, and so many epistemic surprises may await her interaction that is therefore more comparable to human-to-human exchange than to human-to-tool administration.

To summarize the theoretical arguments, the first three stages represent a systematization of key criteria from the extended cognition literature. Yet, the alternative hypothesis of extracted cognition is not directly revealed by them, because the case

of Anna can, but need not, agree to those criteria. Decisive differences only emerge in the additional fourth stage where the issue of personal integration, as further developed, can delineate the Otto and Anna cases. If those three extra criteria point us toward the extracted framework for increasingly more cases of tool interactions, then extended cognition loses some of its grounding. In short: Instead of expanding our reach of cognitive control, perhaps we in fact develop and use increasingly more tools that impose causal, epistemic, and decisional constraints upon our cognitive ways of being, which ultimately results in the dissipation of some of our cognition.

4.2 Practical arguments

Along with the theoretical arguments, we can also arbitrate between extended and extracted cognition in terms of the cognitive practices taking place between agent and tool. For example, Clark (2003) evolves this practice-oriented access throughout, and Menary (2006, pp. 330–331, 2007, Chap. 6, 2010a, pp. 238–241) focuses on it in terms of tool manipulations, their norms, and how they consequently transform our cognitive abilities (also see Hurley, 1998; Rowlands, 1999, 2010; Wilson, 2004; Wheeler, 2005). This argumentative access shows that there is no metaphysical interest in whether tools are themselves cognitive, which externalists never seriously defended nor investigated (e.g., Clark, 2010a, p. 89). To paraphrase Menary on this: 'Tools aren't cognitive, our manipulations of them are.' (cf. Menary, 2010c, p. 617). Sticking to our example scenarios, both Otto and Anna exercise tool practices of providing and receiving cognitive information (e.g., who do I meet next?) and cognitive strategies (e.g., how do I get there?). The practical question is then: In which of two possible directions does a concrete cognitive practice move the strategy or information—toward or away from the agent? After exposing the neutrality of both directions, we can see how purpose and consequences of our practices indicate one or the other kind of cognition.

Cognitive outsourcing describes any cognitive practice that carries information or strategies from central to peripheral vehicles of the broadened cognitive system (or even out of that system). We shall use this here only as a descriptive label, though such practice has of course an important normative dimension as well (cf. Frischmann & Selinger, 2018). Two brief examples: Anna has the sudden idea to meet her father at the MoMA. After a short voice prompt, her AI self-sufficiently writes a message to Otto, waits for the confirmation, and finally enters all meeting information in Anna's calendar. Otto, upon receiving the invitation, also outsources this information as a reminder into his notebook. In both cases, the crucial information is carried away from the center of the cognitive system. Similar examples can be given for the outsourcing of cognitive strategies.

Cognitive insourcing describes any cognitive practice that carries information or strategies from outside the cognitive system into its peripheral, or then even to its central, vehicles. Here are two more examples: Anna's AI assistant may download further apps that, perhaps together with her smart wearables, displace the maturation and manifestation of cognitive skills usually required for the present task—for example, route planning, time managing, and spatial orientation to meet her dad. Otto may instead obtain a city map for his notebook that he uses so often as to render it

another peripheral part of his extended cognitive system. In both cases, the crucial information or strategy is moved from outside to inside the cognitive system (liberally framed).

In these four scenarios, we see how Anna and Otto employ the same general practices (cognitive insourcing and outsourcing) to achieve the same cognitive objectives respectively, yet their individual intentions (teleological perspective) and the consequences (developmental perspective) are so radically different that Otto may have extended and Anna extracted cognition. In case of conflict, the actual consequences shall have a greater impact on the type of cognition we ascribe than the preceding intentions. So how do these two perspectives (purpose and impact) establish the divide between extended and extracted cognition in practical regards?

Otto intends to use his notebook to gain a deeper understanding, to learn and improve certain skills, and so the tool does not replace, but intensifies his involvement. As an extended cognizer, he tries to causally control and epistemically exploit his cognitive tools such that he continues to be in charge, which may eventually earn him cognitive improvements. From a developmental perspective then, repeated tool use is conceived to "turbo-charge" our cognitive abilities (cf. Clark, 2010b, p. 59). So since externalists are primarily motivated by the idea of cognitive enhancements, extended cognition conceives of the cognitive center as augmented and improved by any of the agent's tool practices.

In contrast, Anna is made to use her AI out of mere pragmatic and productive intentions, not to learn and deepen any of her skills, but to give their execution for task completion away so that she can bother less about it. Anna is more invested in the cognitive business of delegation, riddance, or even retirement of her own cognitive duties and skills. Therefore, extracted cognition conceives the cognitive center as reluctant toward a stronger cognitive involvement. Extracted cognition may also be ascribed in cases where the AI is producing outputs for which Anna would usually use a certain skill that she is now prevented from manifesting or even acquiring in the first place. For example, the AI may offer to solve a task by coding something in a programming language Anna never heard of. When her tool does it so fluently and easily for Anna, she is at least momentarily reluctant to really learn and master this skill herself. Perhaps we then only observe a mere shift of learning focus for Anna, because the cognitive replacements by her tool make corresponding internal skills obsolete, and yet others arise and require attention instead. But even if such mild consequences are the norm, it already concedes that many such tool innovations cause cognitive negligence at one point so to enable enhancements at other instances. These cognitive tradeoffs, as mentioned earlier, would need to be evaluated on a case-by-case basis, since some refocus on other, meta-cognitive skills (required for the tool usage by the agent) may be valued much more than the skills we abandon because of this usage. However, in the case of Anna, she initially only learns AI managing skills that are much shallower, highly specialized, and less universally applicable than those she abandons or ignores. More serious however, if no successor skill is learned or trained, perhaps because the AI becomes so advanced, Anna's isolated cognitive capacities and perhaps even capabilities might slowly decline (cf. Heinrichs, 2024)—but this is an empirical possibility to be tested. Different severities arise depending on Anna's skills and their proficiency (if any) before these technolo-

gies obliterated her internal cognitive involvement (Cassinadri, 2024, pp. 5–7, 12). Hence, this developmental perspective rather views such tool practices as 'turbo-*dis*charging' our cognitive skills.

Importantly though, extracted cognition does not demand that a certain skill is fully eliminated or irreversibly hindered from acquisition by the agent, but only that the interaction with a more capable tool is at least temporally blocking the skill to manifest or even mature in the organism. As this shows, not the tools themselves, but the way we implement them into our cognitive practices can have drastic effects on our overall cognitive reality such that either extended or extracted cognition is the better ascription.

4.3 Ethical arguments

While theoretical and practical arguments dominate the debate, some also contrive ethical arguments. These address our moral and legal intuitions for assigning one kind of cognition over the other (Levy, 2007; Carter & Palermos, 2016; Fasoli, 2016; Heersmink, 2017; Heinrichs, 2017, 2024; Hernández-Orallo & Vold, 2019; Drayson & Clark, 2020; Farina & Lavazza, 2022; Cassinadri, 2022). Any such ethical argument is driven by the following question: How do we expect our cognition to be? This approach compares the actual (descriptive) with the desirable (prescriptive) cognitive reality of individuals. The underlying assumption here is that 'cognition' is not a purely descriptive notion that can be solely ascribed based on objective facts about the candidate system. Instead, it is arguably also, at least partly, a prescriptive concept that we apply based on normative expectations about our cognitive standing. The theoretical and practical approaches may be conjoined with this tactic, though the overall dispute can be pursued either way. A crucial ethical argument for extended cognition was hinted at only in passing by Clark and Chalmers (1998, p. 18). After embellishing this argument for Otto, we can reverse its maneuver to argue for extracted cognition in the case of Anna.

Otto heavily depends on his notebook for many of his cognitive tasks, especially when we refocus on his Alzheimer's disease. Now suppose he is robbed of his notebook on his way to the MoMA. Without any (external) memory of the address or route, he cannot employ his orientation skills and will not manage to get there, and many other general abilities are severely impaired or hindered, too. Months later at court, if the notebook is understood as a quite dispensable, non-cognitive accessory, then the judicial decision would be rather harmless for the thieves. But, as Otto's lawyer argues, the theft was rather a severe cognitive infringement given that the notebook is as important as corresponding neural processes in other people. This sentence would however require that Otto is in fact an extended cognizer consisting of organism plus notebook (cf. Carter & Palermos, 2016, pp. 551–552). We can call the argumentative tactic *cognitive blackmailing* (cf. Cassinadri, 2022, p. 7): We shall better consider Otto an extended cognizer so as to give greater protection and moral value to his external cognitive vehicles, because otherwise we risk that such attacks are wrongly belittled—all despite the quite objective harm potential for Otto and parts of his (external) cognition. Here we now see that the ascription of extended cognition to Otto heavily depends on the correctness of his medical report. If he overly

relied on his notebook to outsource his memory *just so*, then perhaps the judge would see this as his own fault. So without this pathological element, Otto could not defend his notebook usage as a necessary surrogate tactic, and this would have robbed Clark and Chalmers the moral thrust for their ascription of extended cognition.

However, the opposite conclusion of cognitive extraction for Anna cannot be defended based on tool loss, but it requires a scenario of technological excess—which might equally cause cognitive poverty for Anna as when Otto is without his notebook. Suppose Anna's AI is automatically instructing her to go see her dad at the MoMA. Her digital ecosystem is starting the navigation: her smart watch vibrates at every intersection and her smart glasses augment an extra layer of reality onto the pavement for her (cf. Clark, 2014, p. 172). She does not even need to look up from her phone a single time until she arrives. Shockingly though, the MoMa she stands in front of does not at all look as expected, because it is in fact a design store in SoHo.[1]

While Otto does not reach the MoMA because he cannot use his tool, Anna does not get there because she is excessively using it. Losing the AI gadgets may be one option to bring her back on track—assuming that this would incentivize Anna to use and train her internal cognitive skills again, which is an option Otto does not have as an Alzheimer's patient. Importantly, Anna's AI system did not suffer from any technical failure, but its routines still depend on some human guidance and interventions that Anna did not care to provide, since she is overly confident in the tool's unchecked suggestions. In this not too unrealistic scenario, we should understand Anna as an extracted cognizer, since she did not actually think with her tool, but she stopped thinking because of it. Anna lets the AI govern her decisions and behaviors to the point of danger. And even if such errors hardly ever occur, exactly this (seemingly) high level of tool performance might create the type of trust by Anna that our cogno-normative standards should reject: No matter your tool's intellectual proficiency next to you, do not stop thinking because of that. This normative intuition can be triggered even stronger with more extreme scenarios, but this one example suffices to reverse the cognitive blackmailing introduced for Otto: We shall better consider Anna in danger of becoming an extracted cognizer so to give greater protection and moral value to her internal cognitive vehicles, because otherwise we risk that such tools extract and cancel parts of them—all despite the quite objective harm potential for Anna and her (internal) cognition.

Though what if we were to imagine Anna as having, for example, some neurodegenerative disease that forces her to create an external backup of her cognitive routines—just like Otto has to do it for his memory? Externalizing those skills would be the only way she can save and maintain those parts of her cognitive system. So as above, this would invert our moral judgment about the kind of cognition we should ascribe to her. In conclusion, what ultimately seems to matter for our moral evaluation is the developmental dimension, what the agent can or cannot do about it, and what we would then expect to be the best tactic and outcome (cf. Heinrichs, 2024). Though as Anna's case highlights, we firstly grant protection to organismic cognitive structures as long as possible (cf. Heinrichs, 2017).

[1] This example is gratefully borrowed from an anecdote by, but not about, Achim Stephan.

5 The future

The presented scenarios and arguments make us reconsider how our present and future interactions with advanced technologies shape the kind of cognition we have and conceive for ourselves. Two hypotheses have been discussed in this context.

Extended cognition generally envisions our tool practices as cognitive augmentations that enhance and empower our overall and individual capabilities in some way—Otto-*con*-notebook can memorize things because Otto-*sin*-notebook has acquired skills for the clever use of his tool. So while individuals employ tools to attain and perhaps increase their individual cognitive autonomy, the mass-installation of cognitive tools shall bring collectives closer toward cognitive democratization: people gain power over their cognitive condition and intellectual outcomes.

Extracted cognition addresses the fact that we are designing AI gadgets which shall learn and emulate our cognitive skills, such that their subsequent task performance substitutes our cognitive involvement, which then ultimately dispenses our own cognitive attendance in such settings—the AI thinks like Anna *for her*, because that is how Anna uses it. So while individuals may fall under intellectual hegemony, collectives may follow a trend of cognitive autocratization: tools gain power over our ways of thinking, or what will be left of it. What feels provocative writing today, may read understated tomorrow. A few potential points of critique against this hypothesis are finally worth defending.

First, some may understand the hypothesis of cognitive extraction as a techno-pessimistic, or a socio-psychological, or even a medico-pathological criticism of current technological advancements and their sedimentation into our societies. These are possible, but different philosophical stances that extracted cognition, on its own, is not committed to (cf. Paglieri, 2024 for a techno-pessimistic stance). The present investigation has only descriptive interests within the philosophy of cognition, namely to better understand the situated nature of our cognizing in virtue of these novel tool interactions. Perhaps many people are able to form interaction strategies with AI tools that enrich and empower their inner and overall cognitive capabilities, but others may be more susceptible to having or letting their cognition be extracted to such tools. And still, only after such cognitive ascriptions are settled can their normative evaluations come into view. This even holds for the presented ethical arguments as they deal with normative intuitions *toward*, but not *about* or *from*, each hypothesis.

Second, the present discussion of extracted cognition exclusively relied on advanced technologies, but the range of applicable scenarios should be more diverse to properly compare extended and extracted cognition as coequal hypotheses. Indeed. Consider this social scenario then: Otto may just as well substitute his notebook, now that it has been stolen, with his partner in life Elle (cf. Clark & Chalmers, 1998, pp. 17–18). Given her excellent biological memory of his notebook, Otto can adapt his memory routines toward Elle instead. In analogy, Anna may appoint Izzi as a new assistant worker who is, just like her laptop, eager to learn her skills and always offers Anna to take over all her tasks—perhaps thereby motivating the company to replace Anna with Izzi down the road. So it seems, tentatively at least, Otto can extend his cognition to Elle, and Anna may extract her cognition to Izzi. With some other examples like this, both extended and extracted cognition are more widely applicable.

Third, can all cognitive skills be extracted until a state of cognitive oblivion is reached? On the tool side, more and more of our cognitive skills can be successfully captured and imitated. On the human side, however, we may reach a practical limit to the extent and kinds of cognitive processes we (let) extract—some cognitive capabilities may belong to our un-extractable core of cognition. But time will tell how low we can go in this game of cognitive limbo.

Fourth, did we not always study and then at times demonize new tool developments and their potential impact on our human condition—so what is so new and important about extracted cognition then? Such a motive does indeed have a long intellectual history, but it primarily amounted to a loose empirical worry, and not to a more precise theoretical position. Hence, extracted cognition is primarily not an empirical thesis over how our cognitive conditions must or will be affected, because it mainly provides a theoretical framework to properly found such investigations to follow. To this end, it reconceives the kind of cognition we have when interacting with tools that we did not, and could not have, seriously considered before. And so it points at a cognitive phenomenon we had no means to seriously think and converse about to this point. By philosophical standards, this is a contribution novel and relevant enough for further theoretical, and perhaps empirical, inquiries.

Fifth, the extended cognition literature already addressed vastly different types of cognitive interactions with varying cognitive impacts. If this were the main and sole issue, we can certainly agree that all such cognitive interactions form a continuum: from full internalization of the tool, over mutual in-the-loop cooperations, to full externalization away from the agent (cf. Hernández-Orallo & Vold, 2019); and from immensely increasing to drastically decreasing our internal cognitive loads (Cassinadri, 2024, p. 2). Otto and Anna shall mark the respective extremes. Agreeably too, we often maintain a multitude of interactions to a single tool (Fasoli, 2018, p. 679)—for example, the innumerable ways we use our smartphones with greatly varying cognitive impacts. Or suppose Otto were healthy, then his notebook usage would greatly reduce his internal cognitive load as he would not (need to) employ his internal memory capacities anymore—the memory structure is now simply external. Perhaps externalists would therefore defend that the usage of advanced AI tools makes no difference to their thesis, because they openly acknowledge that tools greatly simplify our ways of thinking, but this shall not prevent them from becoming constitutive parts of our extended cognitive systems. Hence, why should we now contrarily or additionally contrive the hypothesis of extracted cognition?

This seemingly powerful critique helps us to close the circle. We initially asked what kind of cognition we have. Extended and extracted cognition are the two proposals discussed. Both hypotheses go beyond discussing different kinds of *cognitive interactions* that we (are made to) situate ourselves with. Instead, it is the entirety of an individual's cognitive interactions that allows us to apply our theoretical, practical, and ethical considerations, and to then determine the *kind of cognition* we ascribe to that individual. Extended cognition views us as thinking beings that adapt (to) tools and structures so to solve our cognitive tasks as a joint effort. Extracted cognition views us as thinking beings that exploit tools and structures in order to solve our cognitive tasks with minimal to no internal cognitive engagement. Externalists already pointed at a whole range of different cognitive interactions and impacts, but

the presented arguments explicate why they offer a misleading interpretation of some cases and might not be well-prepared for newly arising ones. In turn, extracted cognition does not only invoke those new extreme cases, but it also offers a revisionary interpretation for some established cases—healthy Otto might be one of them.

From this wide view on cognition, the step toward mind is now much smaller. The extended mind is control-focused and involved. However, the extracted mind is energy-focused and retreated. The spirit and reasons for the latter have been amply presented in this paper, but perhaps it is precisely these current developments in our technological and cognitive lives that make extended theorizing more needed than ever: How can we interact both productively and beneficially with our cognitive artifacts while avoiding our own cognitive clearance? As Clark framed his 'advanced cognition' early on: "We use intelligence to structure our environment so that we can succeed with *less* intelligence. Our brains make the world smart so that we can be dumb in peace!" (1997, p. 180). If things literally turn out this way, this peace may not be of the extended, but of the extracted mind.

Acknowledgements The author would like to thank (in chronological order) members of the philosophical "Oberseminar" of the Institute for Philosophy at Osnabrück University (organized by Nikola Kompa and Susanne Boshammer), guests of the workshop "Central Topics in Situated Cognition" by the Research Training Group "Situated Cognition" at Ruhr University Bochum (organized by Albert Newen, Nikola Kompa, Julia Wolf, and the author), members of the philosophical Reading Club "Affectivity" of the Institute of Cognitive Science at Osnabrück University (organized by Achim Stephan, Sven Walter, Gregor Hörzer, and Imke von Maur), guests of the 49th Annual Philosophy of Science Conference at the Inter-University Centre Dubrovnik (organized by Joseph Berkovitz et al.), guests of the 31st Conference of the European Society for Philosophy and Psychology in Grenoble (organized by Adrian Alsmith et al.), guests of the 46th Annual Meeting of the Cognitive Science Society in Rotterdam (organized by Larissa Samuelson et al.), and three anonymous reviewers that greatly shaped the final paper. The author would also like to thank the German Research Foundation DFG which supported this associated project in the context of funding the Research Training Group "Situated Cognition" (GRK 274877981).

Funding Open Access funding enabled and organized by Projekt DEAL.
Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer GRK 274877981.

Declarations

Competing interests The author declares no competing interests.

References

Aagaard, J. (2020). 4E cognition and the dogma of harmony. *Philosophical Psychology*, *34*(2), 165–181. https://doi.org/10.1080/09515089.2020.1845640

Adams, F. (2010). Why we still need a mark of the cognitive. *Cognitive Systems Research*, *11*(4), 324–331. https://doi.org/10.1016/j.cogsys.2010.03.001

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*(1), 43–64. https://doi.org/10.1080/09515080120033571

Adams, F., & Aizawa, K. (2008a). *The bounds of cognition*. Blackwell. https://doi.org/10.1002/9781444391718

Adams, F., & Aizawa, K. (2008b). Why the mind is still in the head. In P. Robbins, & M. Aydede (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 78–95). Cambridge University Press. https://doi.org/10.1017/CBO9780511816826.005

Adams, F., & Aizawa, K. (2010a). Defending the bounds of cognition. In R. Menary (Ed.), *The Extended Mind* (pp. 67–80). MIT Press. https://doi.org/10.7551/mitpress/9780262014038.003.0004

Adams, F., & Aizawa, K. (2010b). The value of cognitivism in thinking about extended cognition. *Phenomenology and the Cognitive Sciences*, *9*(4), 579–603. https://doi.org/10.1007/s11097-010-9184-9

Adams, F., & Garrison, R. (2013). The mark of the cognitive. *Minds and Machines*, *23*(3), 339–352. https://doi.org/10.1007/s11023-012-9291-1

Aizawa, K. (2010). The coupling-constitution fallacy revisited. *Cognitive Systems Research*, *11*(4), 332–342. https://doi.org/10.1016/j.cogsys.2010.07.001

Andrada, G., Clowes, R. W., & Smart, P. R. (2023). Varieties of transparency: Exploring agency within AI systems. *AI & Society*, *38*(4), 1321–1331. https://doi.org/10.1007/s00146-021-01326-6

Carter, J. A., & Palermos, S. O. (2016). Is having your computer compromised a personal assault? The ethics of extended cognition. *Journal of the American Philosophical Association*, *2*(4), 542–560. https://doi.org/10.1017/apa.2016.28

Cassinadri, G. (2022). Moral reasons not to posit extended cognitive systems: A reply to Farina and Lavazza. *Philosophy & Technology*, *35*(64), 1–20. https://doi.org/10.1007/s13347-022-00560-0

Cassinadri, G. (2024). ChatGPT and the technology-education tension: Applying contextual virtue epistemology to a cognitive artifact. *Philosophy & Technology*, *37*(14), 1–28. https://doi.org/10.1007/s13347-024-00701-7

Clark, A. (1989). *Microcognition: Philosophy, cognitive science, and parallel distributed processing*. MIT Press. https://doi.org/10.7551/mitpress/4597.001.0001

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. MIT Press. https://doi.org/10.7551/mitpress/1552.001.0001

Clark, A. (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. Oxford University Press.

Clark, A. (2007). Curing cognitive hiccups: A defense of the extended mind. *The Journal of Philosophy*, *104*(4), 163–192. https://doi.org/10.5840/jphil2007104426

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195333213.001.0001

Clark, A. (2010a). Coupling, constitution, and the cognitive kind: A reply to Adams and Aizawa. In R. Menary (Ed.), *The Extended Mind* (pp. 81–99). MIT Press. https://doi.org/10.7551/mitpress/9780262014038.003.0005

Clark, A. (2010b). Memento's revenge: The extended mind, extended. In R. Menary (Ed.), *The Extended Mind* (pp. 43–66). MIT Press. https://doi.org/10.7551/mitpress/9780262014038.003.0003

Clark, A. (2010c). Much ado about cognition. *Mind*, *119*(476), 1047–1066. https://doi.org/10.1093/mind/fzr002

Clark, A. (2014). *Mindware: An introduction to the philosophy of cognitive science* (2nd ed.). Oxford University Press.

Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis*, *58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Drayson, Z. (2010). Extended cognition and the metaphysics of mind. *Cognitive Systems Research*, *11*(4), 367–377. https://doi.org/10.1016/j.cogsys.2010.05.002

Drayson, Z., & Clark, A. (2020). Cognitive disability and embodied, extended minds. In A. Cureton, & D. T. Wasserman (Eds.), *The Oxford Handbook of Philosophy and Disability* (pp. 579–597). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190622879.013.10

Farina, M., & Lavazza, A. (2022). Incorporation, transparency and cognitive extension: Why the distinction between embedded and extended might be more important to ethics than to metaphysics. *Philosophy & Technology*, *35*(10), 1–24. https://doi.org/10.1007/s13347-022-00508-4

Fasoli, M. (2016). Neuroethics of cognitive artifacts. In A. Lavazza (Ed.), *Frontiers in Neuroethics: Conceptual and Empirical Advancements* (pp. 63–75). Cambridge Scholars.

Fasoli, M. (2018). Substitutive, complementary and constitutive cognitive artifacts: Developing an interaction-centered approach. *Review of Philosophy and Psychology*, *9*(3), 671–687. https://doi.org/10.1007/s13164-017-0363-2

Frischmann, B., & Selinger, E. (2018). *Re-engineering humanity*. Cambridge University Press. https://doi.org/10.1017/9781316544846

Heersmink, R. (2012). Mind and artifact: A multidimensional matrix for exploring cognition-artifact relations. In J. M. Bishop & Y. J. Erden (Eds.), *Proceedings of the 5th AISB Symposium on Computing and Philosophy* (pp. 48–55). Society for the Study of Artificial Intelligence and Simulation of Behaviour.

Heersmink, R. (2013). A taxonomy of cognitive artifacts: Function, information, and categories. *Review of Philosophy and Psychology*, *4*(3), 465–481. https://doi.org/10.1007/s13164-013-0148-1

Heersmink, R. (2017). Extended mind and cognitive enhancement: Moral aspects of cognitive artifacts. *Phenomenology and the Cognitive Sciences*, *16*(1), 17–32. https://doi.org/10.1007/s11097-015-9448-5

Heinrichs, J. H. (2017). Against strong ethical parity: Situated cognition theses and transcranial brain stimulation. *Frontiers in Human Neuroscience*, *11*(171), 1–13. https://doi.org/10.3389/fnhum.2017.00171

Heinrichs, J. H. (2024). Narrows, detours, and dead ends—how cognitive scaffolds can constrain the mind. In J.-H. Heinrichs, B. Beck, & O. Friedrich (Eds.), *Neuro-ProsthEthics* (Vol. 9, pp. 57–72). Springer. https://doi.org/10.1007/978-3-662-68362-0_4

Hernández-Orallo, J., & Vold, K. (2019). AI extenders: The ethical and societal implications of humans cognitively extended by AI. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 507–513. https://doi.org/10.1145/3306618.3314238

Hurley, S. L. (1998). *Consciousness in action*. Harvard University Press.

Levy, N. (2007). Rethinking neuroethics in the light of the extended mind thesis. *The American Journal of Bioethics*, *7*(9), 3–11. https://doi.org/10.1080/15265160701518466

Menary, R. (2006). Attacking the bounds of cognition. *Philosophical Psychology*, *19*(3), 329–344. https://doi.org/10.1080/09515080600690557

Menary, R. (2007). *Cognitive integration: Mind and cognition unbound*. Palgrave Macmillan. https://doi.org/10.1057/9780230592889

Menary, R. (2010a). Cognitive integration and the extended mind. In R. Menary (Ed.), *The Extended Mind* (pp. 226–243). MIT Press. https://doi.org/10.7551/mitpress/9780262014038.003.0010

Menary, R. (2010b). Introduction: The extended mind in focus. In R. Menary (Ed.), *The Extended Mind* (pp. 1–25). MIT Press. https://doi.org/10.7551/mitpress/9780262014038.003.0001

Menary, R. (2010c). The holy grail of cognitivism: A response to Adams and Aizawa. *Phenomenology and the Cognitive Sciences*, 9(4), 605–618. https://doi.org/10.1007/s11097-010-9185-8

Mühlhoff, R. (2020). Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media sociology of machine learning. *New Media & Society*, *22*(10), 1868–1884. https://doi.org/10.1177/1461444819885334

Newen, A., de Bruin, L., & Gallagher, S. (Eds.). (2018). *The Oxford handbook of 4E cognition*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.001.0001

Norman, D. A. (1991). Cognitive artefacts. In J. M. Carroll (Ed.), *Designing interaction: Psychology at the human-computer interface* (pp. 17–38). Cambridge University Press.

Paglieri, F. (2024). Expropriated minds: On some practical problems of generative AI, beyond our cognitive illusions. *Philosophy & Technology*, *37*(55), 1–30. https://doi.org/10.1007/s13347-024-00743-x

Piredda, G. (2017). The mark of the cognitive and the coupling-constitution fallacy: A defense of the extended mind hypothesis. *Frontiers in Psychology*, *8*(2061), 1–10. https://doi.org/10.3389/fpsyg.2017.02061

Robbins, P., & Aydede, M. (Eds.). (2008). *The Cambridge handbook of situated cognition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511816826

Rockwell, W. T. (2005). *Neither brain nor ghost: A nondualist alternative to the mind-brain identity theory*. MIT Press. https://doi.org/10.7551/mitpress/4910.001.0001

Rowlands, M. J. (1999). *The body in mind: Understanding cognitive processes*. Cambridge University Press. https://doi.org/10.1017/CBO9780511583261

Rowlands, M. J. (2009). Extended cognition and the mark of the cognitive. *Philosophical Psychology*, *22*(1), 1–19. https://doi.org/10.1080/09515080802703620
Rowlands, M. J. (2010). *The new science of the mind: from extended mind to embodied phenomenology*. MIT Press. https://doi.org/10.7551/mitpress/9780262014557.001.0001
Rowlands, M. J., Lau, J., & Deutsch, M. (2020). Externalism about the mind. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Winter 2020 Edition). The Metaphysics Research Lab—Stanford University. https://plato.stanford.edu/archives/win2020/entries/content-externalism/
Rupert, R. D. (2004). Challenges to the hypothesis of extended cognition. *The Journal of Philosophy*, *101*(8), 389–428. https://doi.org/10.5840/jphil2004101826
Rupert, R. D. (2009). *Cognitive systems and the extended mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195379457.001.0001
Shapiro, L. A. (2008). Functionalism and mental boundaries. *Cognitive Systems Research*, *9*(1–2), 5–14. https://doi.org/10.1016/j.cogsys.2007.07.008
Shapiro, L. A. (2019). *Embodied cognition* (2nd ed.). Routledge. https://doi.org/10.4324/9781315180380
Slaby, J. (2016). Mind invasion: Situated affectivity and the corporate life hack. *Frontiers in Psychology*, *7*(266), 1–13. https://doi.org/10.3389/fpsyg.2016.00266
Sprevak, M. (2009). Extended cognition and functionalism. *The Journal of Philosophy*, *106*(9), 503–527. https://doi.org/10.5840/jphil2009106937
Sutton, J. (2010). Exograms and interdisciplinarity: history, the extended mind and the civilizing process. In R. Menary (Ed.), *The Extended Mind* (pp. 189–225). MIT Press. https://doi.org/10.7551/mitpress/9780262014038.003.0009
Varga, S. (2018). Demarcating the realm of cognition. *Journal for General Philosophy of Science*, *49*(3), 435–450. https://doi.org/10.1007/s10838-017-9375-y
Walsh, P. J. (2017). Cognitive extension, enhancement, and the phenomenology of thinking. *Phenomenology and the Cognitive Sciences*, *16*(1), 33–51. https://doi.org/10.1007/s11097-016-9461-3
Walter, S. (2010). Cognitive extension: The parity argument, functionalism, and the mark of the cognitive. *Synthese*, *177*(2), 285–300. https://doi.org/10.1007/s11229-010-9844-x
Walter, S., & Kästner, L. (2012). The where and what of cognition: The untenability of cognitive agnosticism and the limits of the motley crew argument. *Cognitive Systems Research*, *13*(1), 12–23. https://doi.org/10.1016/j.cogsys.2010.10.001
Wheeler, M. (2005). Reconstructing the cognitive world: The next step. MIT Press. https://doi.org/10.7551/mitpress/5824.001.0001
Wheeler, M. (2010). In Defense of extended functionalism. In R. Menary (Ed.), *The Extended Mind* (pp. 245–270). MIT Press. https://doi.org/10.7551/mitpress/9780262014038.003.0011
Wheeler, M. (2011). In search of clarity about parity. *Philosophical Studies*, *152*(3), 417–425. https://doi.org/10.1007/s11098-010-9601-5
Wilson, R. A. (2004). *Boundaries of the mind: The individual in the fragile sciences*. Cambridge University Press. https://doi.org/10.1017/CBO9780511606847
Wilson, R. A., & Clark, A. (2008). How to situate cognition: Letting nature take its course. In M. Aydede & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 55–77). Cambridge University Press. https://doi.org/10.1017/CBO9780511816826.004