Philosophy & Technology (2022) 35:87
https://doi.org/10.1007/s13347-022-00580-w

RESEARCH ARTICLE

Minds in the Metaverse: Extended Cognition Meets Mixed Reality

Paul Smart[1,2]

Received: 3 April 2022 / Accepted: 22 August 2022

Abstract
Examples of extended cognition typically involve the use of technologically low-grade bio-external resources (e.g., the use of pen and paper to solve long multiplication problems). The present paper describes a putative case of extended cognizing based around a technologically advanced mixed reality device, namely, the Microsoft HoloLens. The case is evaluated from the standpoint of a mechanistic perspective. In particular, it is suggested that a combination of organismic (e.g., the human individual) and extra-organismic (e.g., the HoloLens) resources form part of a common mechanism that realizes a *bona fide* cognitive routine. In addition to demonstrating how the theoretical resources of neo-mechanical philosophy might be used to evaluate extended cognitive systems, the present paper illustrates one of the ways in which mixed reality devices, virtual objects (i.e., holograms), and online (Internet-accessible) computational routines might be incorporated into human cognitive processes. This, it is suggested, speaks to the recent interest in mixed/virtual reality technologies across a number of disciplines. It also introduces us to issues that cross-cut disparate fields of philosophical research, such as the philosophy of science and the philosophy of technology.

Keywords Extended cognition · Mechanism · Mechanistic explanation · HoloLens · Virtual reality · Mixed reality

1 Introduction

Recent advances in mobile and portable computing, coupled with advances in photonic technology, have given rise to a number of so-called virtual, mixed, or augmented reality devices. One such device is the Microsoft HoloLens. The HoloLens

✉ Paul Smart
ps02v@ecs.soton.ac.uk

[1] Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK

[2] Instituto de Filosofia da Nova, Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa, Campus de Campolide - Colégio Almada Negreiros, 1099-032 Lisbon, Portugal

Published online: 22 September 2022

is a head-mounted mixed reality device that enables users to interact with three-dimensional virtual objects, called holograms. These holograms are rendered in the local (real-world) environment of the human user, where they can be viewed from multiple angles and manipulated using a combination of bodily movements and voice commands. The presence of such interactive capabilities establishes a point of contact with the notion of active externalism (Clark, 2008; Clark & Chalmers, 1998). In particular, debates about extended cognition and the extended mind (the traditional targets of active externalist theorizing) typically refer to situations in which a human individual interacts with one or more extra-organismic resources, thereby creating a world-involving information processing loop that realizes some of the individual's cognitive processes. In discussing the notion of the extended cognition, for example, Clark (2008) suggests that human brains may come to factor extra-organismic resources deep into their problem-solving routines, thereby "creating hybrid cognitive circuits that are themselves the physical mechanisms underlying specific problem-solving performances" (Clark, 2008, p. 68).

The aim of the present paper is to consider a state-of-affairs in which these "hybrid cognitive circuits" involve the use of a mixed reality device, such as the Microsoft HoloLens. This case is what I will call the HoloFoldit case. The HoloFoldit case features a hypothetical system, called the HoloFoldit system, whose functionality resembles that of a real-world citizen science system, called Foldit (see Cooper et al., 2010).[1] The main point of departure between the (hypothetical) HoloFoldit and (real-world) Foldit systems relates to the inclusion of a mixed reality device (i.e., the HoloLens) in lieu of a conventional desktop computer and "physical" display device. This shift is important, I suggest, because it opens the door to forms of cognitive extension in which mixed reality devices and virtual objects (e.g., holograms) are incorporated into human cognitive routines. Such forms of cognitive extension are important, for they are apt to extend the traditional palette of issues and concerns that animate philosophical debates pertaining to extended cognition and the extended mind. They are also poised to provide a new direction for research that seeks to explore the philosophical significance of virtual reality technologies (Chalmers, 2022; Metzinger, 2018; Turner, 2022). This is particularly important given the recent interest in what has been dubbed the Metaverse—an evolution of the contemporary Internet that is expected to provide users with a range of augmented, mixed, and virtual reality experiences (e.g., Mystakidis, 2022).

The structure of the paper is as follows: The HoloFoldit case is introduced in Section 2. Section 3 then outlines a mechanistic approach to the evaluation of extended cognitive systems. This approach yields a number of criteria that must be met by a putative case of extended cognizing. These criteria relate to the problem of cognitive status (Section 4), which concerns the effort to confirm the cognitive status of a world-involving information processing routine; the problem of constitutive relevance (Section 5), which concerns the effort to identify the constituents of a cognitive mechanism; and the problem of cognitive ownership (Section 6), which concerns the effort to tie a cognitive process (or some other cognitive phenomenon) to

[1] See https://fold.it/ [accessed: 28th March 2022].

a specific human individual. The paper concludes by suggesting that the HoloFoldit case ought to be regarded as a *bona fide* case of extended cognizing. At the very least, none of the criteria associated with a mechanistic approach to extended cognition would seem to negate the idea that mixed reality devices and virtual objects cannot form part of the "hybrid cognitive circuits that are themselves the physical mechanisms underlying specific problem-solving performances" (Clark, 2008, p. 68).

2 The HoloFoldit Case

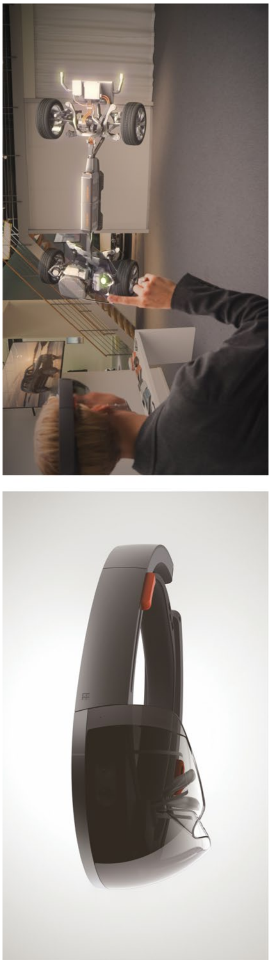
The HoloLens is a head-mounted mixed reality device designed and developed by Microsoft (see Fig. 1a). Unlike conventional virtual reality headsets, such as the HTC VIVE or the Oculus Rift, the HoloLens is a *mixed* reality device. This means that virtual objects (called holograms) are rendered in the local (real-world) environment of the HoloLens user, thereby leading to a mixed (or blended or augmented) reality experience. In essence, the HoloLens enables users to view and interact with real-world physical objects at the same time as they view and interact with virtual holographic objects (see Fig. 1b).

The HoloLens features a so-called inside-out sensor fusion system that supports the spatial mapping of the local physical environment and the tracking of user behavior. These capabilities enable holograms to be rendered at specific locations within the physical world. If, for example, a three-dimensional hologram is rendered in the middle of a room, then it will remain in that position as the user moves around the hologram. This enables the user to view the hologram from multiple perspectives, just as they would a standard physical object.

Users interact with holograms via a so-called Gesture, Gaze, and Voice (GGV) interface. The gaze component of this interface enables holograms to be selected simply by looking at them. Holograms can then be manipulated (e.g., moved, rotated) by implementing specific hand movements (i.e., gestures). (These gestures are detected by forward-facing cameras that are built into the HoloLens device.) Finally, the voice component of the GGV interface enables holograms to be manipulated via the use of voice commands and a speech recognition system. In short, the HoloLens device is able to monitor the activities of the human user and respond to these activities by changing the properties of virtual (holographic) objects.

In addition to its spatial mapping, user tracking, and holographic rendering capabilities, the HoloLens features a WiFi connection that enables it to access online (Internet-accessible) services. While such capabilities are typically used to support the generation of virtual objects, they can also be used to interact with the physical environment via the Internet of Things (IoT). Consider, for example, a state-of-affairs in which the user "flicks" a holographic light switch. Via the IoT, this can be used to dim the lights in the real-world environment, thereby improving the visibility of holographic objects.

To highlight the role of the HoloLens in supporting episodes of extended cognition, it will help to organize the discussion around an example application. The application to be considered here is what I will call the HoloFoldit application. Just



(a) (b)

Fig. 1 The Microsoft HoloLens. **a** The Microsoft HoloLens device (1st generation model). **b** An example of a HoloLens application. In this case, the HoloLens device enables a human user to visualize and interact with a virtual object. (Used with permission from Microsoft) to be clear, this is not a real-world application that can be downloaded and run on the HoloLens device. It is, instead, a fictional (hypothetical) application—one that is "designed" solely for the purpose of guiding the present philosophical analysis.

The functionality of the HoloFoldit application is based on an existing, real-world application, called Foldit (Cooper et al., 2010). The goal of Foldit is to support the derivation of solutions to what is called the protein folding problem, which is a central problem in structural biology (see Dill & MacCallum, 2012). The protein folding problem centers on the attempt to infer (or predict) the three-dimensional structure of a protein molecule based on nothing more than information about the amino acid chain that comprises the molecule. Protein molecules consist of a linear chain of amino acids, which spontaneously fold into a three-dimensional structure as they are being manufactured by the bio-cellular machinery. This folded form of the protein molecule is what is known as the protein's native state or native conformation. The protein folding problem is the problem of predicting the native state of a protein molecule based on information about its (linear) amino acid sequence. This problem is not straightforward. The protein folding problem can be understood as a search through a complex space of conformational possibilities, where candidate conformations are evaluated according to a free energy function (with lower free energy indicating better solution candidates). Computational techniques can be used to explore this space, but most attempts at systematic exploration are rendered intractable due to the complexity of biologically significant protein molecules.

Foldit represents an attempt to solve the protein folding problem by combining human cognitive capabilities with computational search techniques. Users of the Foldit system are presented with a three-dimensional view of a partially folded protein molecule (see Fig. 2). They are then able to effect changes in the protein structure by invoking a number of tools. These include tools to rotate and reconfigure specific regions of the protein molecule. Users are also able to impose soft constraints (called "rubber bands") that limit changes to specific regions of the molecule. Finally, users are able to invoke a number of automatic methods that perform local optimization and reconfiguration tasks. These include a "shake" method (performs combinatorial side-chain rotamer packing), a "wiggle" method (performs gradient-based minimization), and a "rebuild" method (performs fragment insertion) (see Cooper et al., 2010, for details).

What is important to note here is the way in which human cognitive capabilities are being used in combination with computational techniques. In many cases, the human individual is acting so as to constrain the overall search process. The use of rubber bands, for example, can be used to constrain the regions of the search space that are explored by computational methods. Users can also shape the overall trajectory of the search process by persisting with sub-optimal conformations that would otherwise be rejected by fully automated techniques. This is important, since the structure of the fitness landscape (as determined by the aforementioned free energy function) is somewhat rugged, so sub-optimal solution paths must sometimes be pursued in order to escape local optima. Foldit is thus a compelling example of what is sometimes referred to as human-guided search or interactive optimization (Klau et al., 2010). The general idea behind such techniques is that the efficiency of a computational search process can be improved by exploiting human cognitive

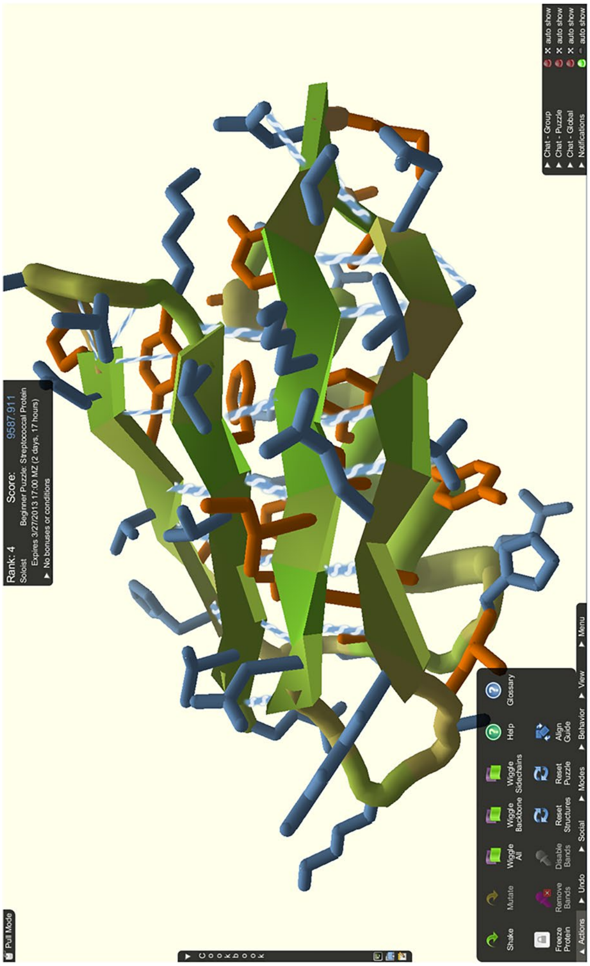


Fig. 2 The Foldit interface. The current state of the protein folding problem is rendered as a three-dimensional view of the protein molecule. The interface provides access to tools that can be used to manipulate the protein structure and add constraints that guide the search for optimal solutions

capabilities. In the case of Foldit, for example, human perceptual and spatial reasoning capabilities are apt to bring a degree of "intelligence" to the search process. Humans are thus able to respond to the presence of visual cues in a manner that might be difficult to detect via computational methods. They are also able to develop innovative solutions to common conformational problems by combining a sequence of user actions into a reusable "recipe" (see Khatib et al., 2011).

In the context of the Foldit system, humans are clearly exerting a great deal of control over the search process. At the same time, however, it would be a mistake to think that humans are doing *all* the heavy-lifting when it comes to the derivation of candidate solutions to the protein folding problem. The Foldit system is, in fact, designed to work as a *hybrid* problem-solving organization, one that solves the protein folding problem by drawing on the complementary capabilities of human individuals and an online suite of computational tools. In order to press maximal benefit from this particular form of bio-technological symbiosis, the Foldit system seeks to support the human user by (in some cases) prompting or cuing the user to engage in certain actions. One example of this comes in the form of visualizations that are intended to highlight specific issues. Consider, for example, that elements of the molecule that need to be internalized (or moved to the interior of the molecular structure) are highlighted as yellow blobs, while atoms that are situated too close together are rendered as red icons.

Now that we have a better understanding of the Foldit system, let us direct our attention to the aforementioned HoloFoldit application.[2] The functionality of the HoloFoldit application closely resembles that of the Foldit system. As with the Foldit system, the user is presented with a three-dimensional view of a partially folded protein molecule. The difference, of course, is that the protein molecule is rendered as a three-dimensional hologram that is anchored to a particular point in the local spatial environment of the user. As with many HoloLens applications, the user is able to interact with this holographic rendering of the protein molecule in a number of ways. They can, for example, rely on the use of voice commands to change their view of the molecule. (Specific examples include the use of "ROTATE LEFT" and "ZOOM IN" instructions.) They can also walk around the hologram (or move towards the hologram), so as to change their view of the molecule. Another visualization strategy involves the user leaning forward so as to position their head inside the holographic protein molecule. This enables the user to view the internal structure of the protein molecule, thereby accessing information that might be unavailable from a more superficial perspective (see Hoffman & Provance, 2017).

In addition to the holographic protein molecule, let us suppose that the HoloFoldit application comes equipped with a holographic toolbar, similar to that seen

[2] As noted above, HoloFoldit is a hypothetical HoloLens application—it is not a real-world application that can be downloaded and installed on the HoloLens device. This is not to say that something like the HoloFoldit application could not be developed for the HoloLens. Indeed, there have been a number of efforts to use the HoloLens for the purpose of visualizing and manipulating molecular structures (e.g., Hoffman & Provance, 2017; Müller et al., 2018), and protein structure visualization remains an active area of research interest for virtual reality enthusiasts (e.g., Cassidy et al., 2020).

in the case of the Foldit application (see Fig. 2). This toolbar provides access to routines that allow the human user to add functionally significant graphic elements to the protein molecule. The user could, for example, select a particular region of the molecule by directing their gaze to the target region. (That is to say, they could simply look at a particular region of the protein molecule.) They could then deploy a hand gesture to add “rubber bands” to the molecule. Finally, they could reposition molecular elements by using voice commands. Similar combinations of gaze, gesture, and voice could then be used to invoke computational routines similar to the shake, wiggle, and rebuild routines available in Foldit. Indeed, insofar as these routines are implemented as online, Internet-accessible services, it is perfectly possible for the HoloFoldit application to exploit the *same* routines as those used by its real-world counterpart. The HoloLens, recall, features a WiFi connection, so there is no reason why the HoloFoldit application could not exploit the same library of online computational routines as those used by the Foldit application.

By using the HoloFoldit application, a human individual is able to participate in the effort to solve protein folding problems. While using the application, the individual is involved in a specific problem-solving process, which I will dub the *Protein Structure Prediction (PSP) process*. This process is, I suggest, a *bona fide* form of extended cognizing. It is, in particular, a form of extended cognition in which both the physical HoloLens device and the virtual holographic objects (e.g., the protein molecule) are incorporated into a cognitive process that we ought to see (in the sense to be developed below) as belonging to the human individual.

While the PSP process is deemed to belong to the human individual, it is arguably not the human individual who is solely responsible for the performance of this process. Instead, the process is being performed by a larger systemic organization that includes the human individual, the HoloLens, the holograms, and a suite of online (Internet-accessible) computational routines. For the sake of convenience, let us call this system the *HoloFoldit system*. Inasmuch as the PSP process qualifies as a *bona fide* form of extended cognizing, then the HoloFoldit system will qualify as an extended cognitive system.

In summary, then, the claim is that the PSP process counts as an extended cognitive process. Subsequent sections of the paper will seek to evaluate this claim. In the next section, I outline an approach to the evaluation of extended cognitive systems that draws on the theoretical resources of neo-mechanical philosophy. This approach is inspired by the purported relevance of mechanistic concepts to the individuation of extended cognitive systems and our understanding of extended cognitive phenomena (van Eck, 2019; Fazekas, 2013; Kaplan, 2012; Smart, in press).

3 Extended Cognition and Extended Mechanisms

Within the philosophical literature, debates about extended cognition and the extended mind are frequently accompanied by an appeal to mechanism-related concepts. In his book, *Supersizing the Mind*, for example, Andy Clark introduces us to the notion of cognitive extension by making an explicit appeal to mechanisms. In contrast to a purely brain-bound or intra-cranial view of cognition, Clark advocates an extended view of cognition. According to this view:

> […] the actual local operations that realize certain forms of human cognizing include inextricable tangles of feedback, feedforward, and feed-around loops: loops that promiscuously criss-cross the boundaries of brain, body, and world. The local *mechanisms* of mind, if this is correct, are not all in the head. Cognition leaks out into body and world. (Clark, 2008, p. xxviii, emphasis added)

Since the publication of *Supersizing the Mind*, mechanistic considerations have started to move to the forefront of active externalist debates (van Eck & de Jong, 2016; Fazekas, 2013; Goldstone & Theiner, 2017; Kaplan, 2012; Miłkowski et al., 2018; Smart, in press; Zednik, 2011). This parallels a recent burgeoning of interest in the philosophical study of mechanisms (Glennan, 2017; Glennan & Illari, 2018a), typically as part of efforts to formulate a theory of mechanistic explanation. According to current thinking, mechanistic explanations are deemed to be of widespread importance in a number of scientific disciplines, including neuroscience, cognitive science, and social science (Bechtel & Abrahamsen, 2005; Craver, 2007b; Craver & Tabery, 2016; Hedström & Ylikoski, 2010; Povich, in press). The general idea is that mechanistic explanations explain phenomena by describing the mechanisms responsible for those phenomena. The thing to be explained here is the phenomenon, which has been conceptualized as a form of object-involving occurrent (Kaiser & Krickel, 2017).[3] The goal of mechanistic explanation is to describe the mechanism that is deemed to be responsible for this phenomenon (i.e., the explanandum phenomenon). According to Glennan (2017, p. 17), a "mechanism for a phenomenon consists of entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon." Mechanisms are thus deemed to consist of entities and activities, which, together, are referred to as the *components* of a mechanism.

Two kinds of mechanistic explanations have been described by philosophers: etiological and constitutive mechanistic explanations (e.g., Kaiser & Krickel, 2017). In etiological mechanistic explanations, the aim is to identify the mechanisms that are causally related to a phenomenon. In this case, the components are related to the explanandum phenomenon via the notion of *causal relevance* (Craver, 2007a). We thus say that some part of the material world (e.g., some object) is causally relevant to an explanandum phenomenon. This differs from the state-of-affairs seen in the case of constitutive mechanistic explanations. Here, the relationship between component and phenomenon is one of *constitutive relevance* (Craver, 2007a), which is a

[3] As noted by Krickel (2017), there is some ambiguity surrounding the notion of an explanandum phenomenon in contemporary theories of mechanistic explanation. In particular, it is unclear whether the phenomenon refers to the behaving mechanism, the behavior of the mechanism, or the behavior of a system that contains the mechanism. For present purposes, I will regard phenomena as object-involving occurrents, as per the analysis of Kaiser and Krickel (2017). A phenomenon is thus an event, state, or process (an occurrent) that is associated with some entity, object, or system (the object).

form of non-causal dependency relationship (see Craver et al., 2021).[4] In the case of constitutive mechanistic explanations, then, it is generally viewed as a mistake to see the explanandum phenomenon as being *caused by* a particular mechanism (or the components thereof). Instead, the relationship between mechanism and phenomenon is one of mechanistic constitution (Baumgartner et al., 2020) or (perhaps) mechanistic realization (Wilson & Craver, 2007).[5]

By combining theories of mechanistic explanation with the more general theoretical resources of neo-mechanical philosophy, we can begin to formulate a mechanistic approach to extended cognition. At the heart of this approach is the idea that extended cognitive phenomena are cognitive phenomena (e.g., cognitive processes) that are constituted/realized by *extended cognitive mechanisms*.[6] That is to say, an extended cognitive phenomenon can be distinguished from a non-extended cognitive phenomenon by appealing to the mechanism that constitutes/realizes the phenomenon. If this mechanism qualifies as an extended cognitive mechanism, then the extended status of the cognitive phenomenon will be confirmed. An understanding of what makes a mechanism a member of the class of extended cognitive mechanisms thus promises to support the effort to identify extended cognitive phenomena.

According to the approach to be adopted here, the class of extended cognitive mechanisms is cast as a proper subset of the class of cognitive mechanisms. A cognitive mechanism is, in turn, defined as a mechanism that is deemed to be responsible (in a constitutive sense) for a cognitive phenomenon. In other words, any mechanism that constitutes/realizes a cognitive phenomenon will qualify as a cognitive mechanism. For the purposes of this paper, I will assume that cognitive phenomena (e.g., cognitive processes) are individuated according to functional criteria,[7] which is to say that we recognize something as a cognitive mechanism based on the nature of the phenomenon that is realized by the mechanism. We do not, I suggest, individuate cognitive mechanisms based on (e.g.) the structural or material properties of mechanisms.

In addition to being cognitive mechanisms, extended cognitive mechanisms are represented as a subset of the class of extended mechanisms. What it means for something to count as an extended mechanism remains a little unclear, since the notion of an extended mechanism has not been the subject of detailed philosophical scrutiny. For present purposes, however, I will assume that an extended mechanism is a mechanism that transcends some sort of border or boundary. In conventional cases of extended cognition,

[4] The distinction between causal and constitutive relevance parallels the distinction between embedded/extended cognition, with embedded cognition relying on the notion of causal relevance and extended cognition relying on the notion of constitutive relevance. For this reason, our primary focus in the present paper will be on constitutive mechanistic explanations.

[5] For the purposes of this paper, I will assume that the terms "mechanistic constitution" and "mechanistic realization" are semantically equivalent.

[6] The mechanisms responsible for extended cognitive phenomena are sometimes referred to as extended (Clark, 2011; Kaplan, 2012; Smart, in press; Zednik, 2011), wide (Miłkowski et al., 2018), or supersized (Clark, 2008) mechanisms. For the sake of convenience, I will refer to these mechanisms as extended cognitive mechanisms.

[7] This is broadly consistent with the way that cognitive phenomena have been conceptualized in at least some parts of the philosophical literature. As noted by Illari and Williamson (2012, p. 131), "[i]n psychology in particular, capacities like memory are often given a purely functional description. There are indefinitely many ways the human brain could divide up the task of remembering things."

the nature of this boundary is relatively clear: it is the biological boundary of the human individual—the borders of skin and skull. Given that this form of extended cognition (i.e., human-centered extended cognition) is the form of cognitive extension that interests us in the present paper,[8] we can define an extended cognitive mechanism as a cognitive mechanism that includes a single human individual (an organismic resource) and one or more objects that lie external to the human individual (i.e., extra-organismic resources). Together, the human individual and the extra-organismic resources qualify as components of the mechanism, which is to say they are constitutively relevant to the cognitive phenomenon that is the focus of our (mechanistically oriented) explanatory efforts.

Our approach thus far can be summarized as follows: In order to determine that we confront a *bona fide* case of extended cognizing, we should first identify the phenomenon *P* we seek to explain and then provide a constitutive mechanistic explanation of *P*. This explanation will explain *P* by describing the mechanism *M* that is responsible (in a constitutive sense) for *P*. If we determine that *P* qualifies as a cognitive phenomenon, then *M* will be cast as a cognitive mechanism. If, in addition, we discover that *M* qualifies as a form of boundary-transcending mechanism (i.e., the components of *M* include a human individual and resources external to the human individual), then *M* will be cast as an extended cognitive mechanism. In this case, we will conclude that *P* qualifies as an extended cognitive phenomenon.

According to this account, cases of (human-centered) extended cognition must satisfy the following criteria:

1. **Cognitive Status Criterion:** The phenomenon *P* that is explained by a constitutive mechanistic explanation should qualify as a phenomenon of the cognitive variety. (According to the above account, this means that the mechanism *M* that constitutes/realizes *P* will qualify as a cognitive mechanism).
2. **Human Component Criterion:** A human individual (or some part thereof) *H* should be a component of *M*.[9]

[8] Other forms of cognitive extension include those centered on plants (Parise et al., 2020), spiders (Japyassú & Laland, 2017), and artificial intelligence (AI) systems (Smart, 2018).

[9] The need for the human component criterion is, I think, less clear-cut than is the case for the extra-organismic criterion. For present purposes, I will assume that this criterion is required. To help us understand the need for the criterion, consider the process of solving long multiplication problems. When these problems are solved in the head, we are content, I assume, to accept that the human individual is engaged in a form of mathematical cognizing. When the individual resorts to pen and paper to solve these problems, however, they are arguably involved in a form of extended cognizing (see, for example, Wheeler, 2010). This is because one or more extra-organismic resources are being used as part of the long multiplication process, and the individual plays an active role in creating and coordinating the flow of information between these resources. In such cases, it appears likely that the human individual is well-placed to serve as a component in the mechanism that is responsible for the long multiplication process. Now compare this with a situation in which a human individual solves long multiplication problems by resorting to the use of an electronic calculator. In this case, the human individual is no longer a component in the long multiplication process, since all the multiplicative activity is being performed by the calculator. Here, the extended status of the mathematical routine looks to be in question (see Roberts, 2012b). One reason for this may be that the human individual is no longer poised to serve as a component in the mechanism that realizes the multiplicative routine. Such cases may be better labeled as instances of "cognitive offloading" (or "cognitive outsourcing"), with the label "extended cognition" reserved for those cases in which a human individual (or some part thereof) is a *bona fide* component in the mechanism that realizes a cognitive routine.

3. **Extra-Organismic Component Criterion:** One of the components of *M* should be an extra-organismic resource *O*. That is to say, *O* should be a component that lies outwith the biological borders of *H*.

These criteria can be mapped to two problems that have been discussed in the philosophical literature. The first of these is what I will call the *problem of cognitive status*. This is the problem of determining whether or not some phenomenon (e.g., a process) ought to be regarded as a *bona fide* cognitive phenomenon. A solution to this problem is required if we are to make any progress with respect to the cognitive status criterion. Within philosophical circles, the problem of cognitive status is typically framed as an attempt to identify the characteristic features of cognitive phenomena (e.g., the things that make a particular process a specifically cognitive process). This is sometimes referred to as the search for a "mark of the cognitive" (Adams, 2010; Adams & Aizawa, 2008; Adams & Garrison, 2013; Rowlands, 2009).

The second and third criteria relate to the componential status of some object. In essence, we need to be sure that the human individual and extra-organismic resources are constitutively relevant to the phenomenon to be explained. This establishes a point of contact with what is called the *problem of constitutive relevance* (Craver, 2007a; 2007b). As noted by Craver (2007a, p. 6), "[t]he problem of constitutive relevance is the problem of saying which parts are components in a mechanism and which are not." Accordingly, a solution to the problem of constitutive relevance promises to provide a means by which we can evaluate the human component criterion and the extra-organismic component criterion.

We thus have two problems to tackle—the problem of cognitive status and the problem of constitutive relevance. Both these problems have been the subject of philosophical debates in the philosophy of mind and the philosophy of science. There is, however, a third problem that needs to be tackled. This is what I will call the *problem of cognitive ownership*. To help us understand the nature of this problem, consider that a human individual may be a component in any number of mechanisms. They may, for example, be a component in a socio-technical mechanism, and such a mechanism will include components (e.g., other individuals) that lie external to the biological borders of the individual. There is, in addition, no reason why the relevant socio-technical mechanism could not support the realization of cognitive phenomena and thus qualify as a (distributed) cognitive mechanism. In this case, we have a cognitive mechanism that includes components that lie outwith the biological borders of a given human individual. Nevertheless, it is not clear that we still confront a case of extended cognizing. This is because the cognitive performances of the larger, socio-technical system are not ones that are typically ascribed to a specific human individual. Rather than explaining the cognitive phenomena assigned to a given individual, we are instead explaining the cognitive phenomena assigned to a larger systemic organization, namely, the socio-technical system. In such cases, we could very well have a robust solution to the problem of cognitive status (we can be sure that the target phenomenon is a cognitive phenomenon) and the problem of constitutive relevance (we know what all the components of the mechanism are). Nevertheless, we will have failed to say why we think the target phenomenon ought to be ascribed to a specific (human) individual and thus why we confront a (boundary-transcending) extended mechanism as opposed to a plain old non-extended mechanism. This, in a nutshell, is the problem of cognitive ownership.

The reason this problem arises is because there is nothing particularly special about the idea of mechanisms that include (as components) a human individual and resources that lie external to the biological borders of that individual. A social mechanism, for example, is likely to be constituted by multiple human individuals (as well as multiple non-human artifacts), but it is hard to see why the phenomenon that is realized by such a mechanism (e.g., a social process) ought to be regarded as an *extended* phenomenon (ditto for the mechanism that realizes the phenomenon). In general, what we are looking for in cases of extended cognition is a state-of-affairs in which some cognitive phenomenon (e.g., a cognitive process) is attributed to a particular cognitive agent. Then, on closer inspection, we discover that this phenomenon is constituted by a mechanism whose components lie beyond the borders of the cognitive agent. In situations where the components do *not* lie beyond these (agent-related) borders, then we will not have a case of extended cognizing.

The problem of cognitive ownership is a recognized problem in the active externalist literature (e.g., Roberts, 2012a, b; Rowlands, 2009; Tollefsen, 2006; Wilson, 2004). As noted by Clark (2011) much of the work relating to extended cognition and the extended mind is:

> […] best seen as an investigation of […] conditions which must be met so as to ensure the *proper ownership* of some candidate extended process by a distinct cognitive agent […] (Clark, 2011, p. 454, original emphasis)

Precisely how this ownership-related problem is to be resolved remains unclear. Existing efforts have appealed to the role of human individuals in creating extended mechanisms (Clark, 2008), controlling the flow of information within an extended mechanism (Wilson, 2004), or as being responsible for the outputs of a cognitive routine (Roberts, 2012a, 2012b). For present purposes, let us simply accept that a mechanistic account of extended cognition will need to tackle the problem of cognitive ownership. By accepting this problem, we encounter an additional criterion that needs to be satisfied by putative cases of extended cognizing. This is what I will call the ownership criterion:

4. **Ownership Criterion**: P is deemed to belong to H or be owned by H. (This may be because M is brought into existence as a result of H's actions [i.e., H is responsible for the instantiation for M] or because H serves as a locus of control and coordination relative to the time-variant structural organization and/or processing dynamics of M.)

To recap the account on offer: The claim is that we ought to evaluate putative cases of cognitive extension by examining the mechanisms that are deemed to be responsible (in a constitutive sense) for cognitive phenomena.[10] Such mechanisms

[10] It should be noted that a distinction is sometimes drawn between extended cognition and the extended mind (e.g., Wheeler, 2019). Arguments for the extended mind typically direct their attention to folk psychological kinds, such as states of dispositional belief. Arguments for extended cognition, by contrast, are typically directed to occurrent phenomena (e.g., processes) that fall within the remit of contemporary cognitive science (e.g., extended memory and extended problem-solving). The present account is specifically directed at extended cognition, not the extended mind. This is not to say that the present account cannot be applied to cases of cognitive extension centered around folk psychological (dispositional) kinds. For present purposes, however, the scope of the present paper is limited to the realm of extended cognitive phenomena (specifically, extended cognitive processes).

will feature as part of constitutive mechanistic explanations that seek to explain phenomena by describing the mechanisms responsible for those phenomena. By itself, however, this will not reveal the presence of extended cognitive phenomena, since the strategy of mechanistic explanation is common to both extended and non-extended (e.g., brain-based) cognitive phenomena. Accordingly, we need some means of distinguishing between extended and non-extended cognitive mechanisms. For human-centered forms of extended cognition, I have suggested that we encounter an extended cognitive mechanism when we have a mechanism that qualifies as a cognitive mechanism; the components of the mechanism include a human individual and resources external to the individual; and, finally, the phenomenon realized by the mechanism is seen to "belong to" the human individual.

We are now in a position to apply this mechanistic account to the HoloFoldit case presented in Section 2. Inasmuch as the HoloFoldit case is to be regarded as a *bona fide* form of extended cognizing, then the PSP process must qualify as a *bona fide* cognitive process. In addition, the human individual + extra-organismic resources (e.g., the HoloLens device) must be constitutively relevant to the PSP process (and thus components in the mechanism that realizes the process). Finally, in respect of the issue of cognitive ownership, we have to find some means of tying the PSP process to the (biologically bounded) human individual. That is to say, we have to find a means of substantiating the claim that the PSP process ought to be seen to belong to the human individual. The application of these criteria to the HoloFoldit case will be explored in the following sections.

4 The Problem of Cognitive Status

The problem of cognitive status is the problem of determining whether an explanandum phenomenon ought to be regarded as a *cognitive* phenomenon. In the HoloFoldit case, the explanandum phenomenon is the PSP process. As such, the issue to be addressed is whether the PSP process ought to be regarded as a *bona fide* cognitive process.

The first thing to say here is that the purpose of the PSP process is to solve a particular problem, i.e., the protein folding problem (Dill & MacCallum, 2012). As such, it seems reasonable to regard the PSP process as a form of problem-solving process, and problem-solving processes are a category of processes that are typically seen to fall within the empirical remit of contemporary cognitive science (see Kirsh, 2009).

It should also be noted that the PSP process is a form of *search* process. In effect, the aim of the PSP process is to explore a complex space of conformational possibilities with the aim of discovering a globally optimal solution (i.e., the native state of a protein molecule). The protein folding problem is hard, because the complexity of the relevant solution space precludes the use of exhaustive search methods for all but the simplest protein molecules. In this sense, the PSP process is like many forms of problem-solving activity: it is a form of knowledge-guided search through a complex space of solution possibilities.

The search-like nature of the PSP process is interesting given the way that cognitive scientists have approached the study of human cognitive processes. This approach dates back to some of the earliest work within cognitive science, where human problem-solving was characterized as a form of search activity (Newell & Simon, 1976; Simon, 1996). In a more general sense, search has been viewed as a fundamental feature of cognition. Newell (1990), for example, suggests that:

> Search is fundamental for intelligent behavior. It is not just another method or cognitive mechanism, but a fundamental process. If there is anything that AI has contributed to our understanding of intelligence, it is discovering that search is not just one method among many that might be used to attain ends but is the most fundamental method of all. (Newell, 1990, p. 96)

In more recent work, search has been invoked as a common framework for understanding many aspects of cognition (Fu et al., 2015; Hills et al., 2015). Fu et al. (2015, p. 384), for example, suggest that "[s]earch can be found in almost every cognitive activity, ranging across vision, memory retrieval, problem solving, decision making, foraging, and social interaction."

From a cognitive science perspective, then, there appear to be a number of reasons to accept the cognitive status of the PSP process: The PSP process, as it occurs in both the Foldit and HoloFoldit systems, is a form of problem-solving activity, and this activity relies on a degree of intelligence to guide the search for candidate solutions.[11]

Another way of approaching the problem of cognitive status is to appeal to what has become known as the *parity principle*. According to this 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 [it] as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process. (Clark & Chalmers, 1998, p. 8, original emphasis)

As noted by Clark (2011), the parity principle is best seen as a heuristic for evaluating putative cases of cognitive extension. In particular, the idea behind the parity principle is:

> […] to invite the reader to judge various potential cognitive extensions behind a kind of 'veil of metabolic ignorance'. A good way to do this is ask yourself, concerning some candidate cognitive process P, whether if you were to find P (or better, its functional equivalent) occurring inside the head of some alien organism, you would tend to class P as a cognitive process? (Clark, 2011, p. 449)

In the HoloFoldit case, the "candidate cognitive process" is the PSP process. Accordingly, what we are being asked to do is imagine a state-of-affairs

[11] An issue that may be worthy of further exploration is the *predictive* nature of the PSP process—the idea that we are trying to predict the native conformational structure of a protein molecule based on its linear amino acid sequence. The predictive nature of the protein folding task establishes a point of contact with recent attempts to tie the mark of the cognitive to issues of predictive processing and the minimization of free energy (see Kersten, 2022; Kiverstein & Sims, 2021).

in which the PSP process is being performed inside the head of a given individual (i.e., without the use of bio-external props, aids, and artifacts). If we are inclined to accept the cognitive status of the PSP process when it is performed inside the head of an individual, then the claim is that we ought to accept the cognitive status of the PSP process when it is subject to an alternative form of mechanistic realization. The question, then, is whether we would be inclined to regard the PSP process as a *bona fide* cognitive process if it were to be performed inside the head of a human individual (or, as Clark notes, inside the head of some alien organism).

Individual readers may, of course, differ in their response to this question. Personally, however, I can see no reason why we would fail to regard the PSP process as a cognitive process if it were to be performed inside the head of a human individual. The PSP process is clearly a problem-solving process of sorts, and it is one that relies on the spatial reasoning capabilities of a human subject. This is, after all, the reason why humans are being "incorporated" into the PSP process in the first place. If it were possible to rely on "brute force" computational search techniques to solve problem folding problems, then there would be little reason to involve human agents in the protein folding task—we could just rely on exhaustive search techniques (i.e., techniques that did not require any form of intelligent guiding). This is not to say that protein folding problems are not, on occasion, solved by such (exhaustive search) techniques, or that, in the future, such problems may be solved in a manner that does not require human input (such problems may, for example, be solved by quantum computers using techniques that are akin to an exhaustive search of the solution space). Crucially, however, these are not the sorts of processes that we are being asked to imagine as part of our attempt to apply the parity principle to the HoloFoldit case. What we are being asked to consider is the PSP process as it occurs in the context of the HoloFoldit system. Such a process ought to be counted as a cognitive process, I suggest, on the grounds that if it were to be performed inside the head of a human individual then we would have no problem in accepting it as a cognitive process.

In response to the problem of cognitive status, then, I suggest that we ought to regard the PSP process as a *bona fide* cognitive phenomenon (whether it counts as an *extended* cognitive phenomenon is, of course, a different matter). The reason for this is that if the process were to be performed inside the head of a human individual, then we would have no problem in accepting the process as a cognitive process. If, for example, we were to encounter a human individual that was able to solve protein folding problems in-the-head, wouldn't we be happy to conclude that the individual was engaged in a form of cognitive activity? And wouldn't we also be inclined to acknowledge the successful completion of the protein folding task as a form of cognitive achievement? If the answer to these questions is "yes," then why should we reject the cognitive status of the PSP process simply because it is subject to an alternative form of mechanistic realization—a form of realization that is not wholly dependent on the whirrings and grindings of the biological brain?

5 The Problem of Constitutive Relevance

Having tackled the problem of cognitive status, we now need to confront the problem of constitutive relevance. The problem of constitutive relevance, recall, is the problem of individuating the components of a mechanism. The inter-operation of the various components in a mechanism constitutes/realizes the explanandum phenomenon, which, in the present case, is the PSP process. Thus, if the thing we want to explain is the PSP process, then we need to identify the components of the mechanism that constitutes/realizes this process. This is relevant to the notion of extended cognition, since (according to the account presented in Section 3) extended cognitive processes are realized by extended mechanisms, and what makes something an extended mechanism is the fact that it is a form of boundary-transcending mechanism. What this means, in the specific context of the HoloFoldit case, is that we are looking for a mechanism whose components include the human individual, the HoloLens device, the virtual objects (i.e., holograms), and the online computational systems. These resources should be constitutively relevant to the explanandum phenomenon, which is to say that they must form part of a common, integrated mechanism that is responsible for the PSP process.

The question, of course, is how do we resolve the problem of constitutive relevance? How, in short, do we determine that each of the aforementioned resources is constitutively relevant to the target process?

In response to this question, philosophers have sought to provide an epistemic account of constitutive relevance, i.e., an account that "lists the criteria by which scientists can identify the components of mechanisms in empirical practice" (Prychitko, 2021, p. 1829). The most popular of these accounts is what is known as the Mutual Manipulability (MM) account (Craver 2007a, 2007b). According to this account, issues of constitutive relevance are resolved by experimental interventions that reveal the relationship between component-level and phenomenon-level variables. Given the emphasis assigned to experimental interventions, this is what is sometimes dubbed an interventionist approach to the problem of constitutive relevance (e.g., Kaplan, 2012).

A recent extension to the MM account comes in the form of the Matched Interlevel Experiments (MIE) account (Craver et al., 2021). This account was developed to address problems arising from the original MM account (see Baumgartner & Casini, 2017; Baumgartner & Gebharter, 2016). According to Craver et al. (2021), constitutive relevance is to be understood as a form of "causal betweenness" (see also Harinen, 2018; Prychitko, 2021). What this means is that a component must form part of a causal path that connects two events that together delimit the temporal bounds of the explanandum phenomenon (i.e., the events that mark the beginning and ending of the explanandum phenomenon). In short, a putative component (X's ϕ-ing) is constitutively relevant to an explanandum phenomenon (S's ψ-ing) if X's ϕ-ing lies on a causal path between two variables, namely, an input variable (ψ_{IN}) and an output variable (ψ_{OUT}), where ψ_{IN} represents the input to S's ψ-ing and ψ_{OUT} represents the output of S's ψ-ing. Applying this to the HoloFoldit case yields the following variable assignments:

- S = the HoloFoldit system;
- ψ = the PSP process (the process performed by the HoloFoldit system);
- X = an entity (e.g., the HoloLens device or the human individual);
- ϕ = the activity of X;
- ψ_{IN} = the unfolded (or partially folded) protein molecule (the trigger for S's ψ-ing); and
- ψ_{OUT} = the fully folded protein molecule (the output of S's ψ-ing).

The details of the MIE account need not concern us here. For present purposes, what is important is simply the idea that constitutive relevance can be understood as a form of causal betweenness, and that this "betweenness" can be revealed by certain types of experimental intervention.

The question, then, is whether the individual human agent, the HoloLens device, and the holograms all lie on a causal path that connects ψ_{IN} and ψ_{OUT}. I suspect that many readers will be content to accept that this is, indeed, the case. After all, relative to the way the case is described in Section 2, the claim that there is an ongoing causal exchange between the human individual and the HoloLens device ought to be largely uncontroversial: the HoloLens is continuously tracking the movements of the human user, and its activity changes in response to these movements. The activity of one putative component (i.e., the human individual) is thus exerting a causal influence on another putative component (i.e., the HoloLens). The HoloLens is, in turn, causally responsible for the current state of the holographic projection, and the holographic projection exerts a causal influence on the human individual. In addition to this causal loop (human → HoloLens → hologram → human), there is a further causal loop that runs from the HoloLens device out into the online realm: At certain points, the HoloLens invokes the execution of online computational routines, and these routines deliver results that causally affect the activity of the HoloLens device. The upshot is that the human individual, the HoloLens, the holograms, and the remotely situated computational resources, all form part of a causally interacting nexus that supports the transformation of an initial problem state (a linear amino acid chain) into a problem solution (the fully folded problem molecule). It seems then that all the putative components must be constitutively relevant to the PSP process. They must be constitutively relevant, since all of the aforementioned causal exchanges occur as *part of* the PSP process. That is to say, they occur in response to the initiation of the PSP process (ψ_{IN}), and they all form part of a causal path (or causal chain) that terminates in the solution to the protein folding problem (i.e., the fully folded protein molecule) (ψ_{OUT}). Relative to the notion of causal betweenness, then, the problem of constitutive relevance seems to have been resolved. The human individual and the extra-organismic resources are all constitutively relevant to the target explanandum phenomenon, and they thus form part of the mechanism that constitutes/realizes this phenomenon.

While this is, I think, the correct response to the problem of constitutive relevance, there is an important objection that needs to be tackled before proceeding. This objection relates to the fact that the MIE account relies on the use of experimental interventions to resolve the problem of constitutive relevance. In the HoloFoldit case, however, it should be clear that these interventions have *not* been performed.

The HoloFoldit system, recall, is a purely *hypothetical* system, so it is difficult to see how we could anticipate the results of the sorts of experimental interventions that, according to the MIE account, underwrite claims of constitutive relevance.

In response to this worry, it is important to note that the MIE account is presented as a *sufficient* account of constitutive relevance. The overarching objective of the MIE account is to tell us why the results of certain experimental interventions are sufficient for constitutional claims. This, however, does not mean that constitutional claims are impossible to resolve in the absence of experimental interventions. The point of experimental interventions, recall, is to establish causal betweenness—to confirm that a putative component lies on a causal path between ψ_{IN} and ψ_{OUT}. But this doesn't really tell us anything about the use of alternative means of establishing causal betweenness; it simply highlights the way in which causal betweenness can be revealed via the use of carefully deployed experimental procedures. This is important, for if there were to be some other way of establishing causal betweenness, then there would be little point in performing the experimental interventions—the interventions would, in effect, be rendered unnecessary.

To my mind, this is precisely the sort of situation we confront in the HoloFoldit case. To be sure, the HoloFoldit system is not a real-world system, and thus we cannot subject it to experimental scrutiny. Suppose, however, that such a system were to be implemented in precisely the manner suggested by Section 2. In this case, wouldn't the real-world causal connections simply reflect those that have already been described as part of the HoloFoldit case? And if this is so, then don't we already know what the causal structure of the real-world HoloFoldit system is? We could, of course, subject the real-world HoloFoldit system to experimental scrutiny, as per the MIE account. But what would the results of such interventions really tell us? Would they refute the idea that the HoloLens responds to human gestures, or that the human responds to the current state of the holographic molecule, or that there is a causal loop between the HoloLens and the remotely situated online services? And what impact would the results of such experiments have on our understanding of how all these forms of causal commerce are related to the evolving state of the protein molecule and (when the process concludes) the properties of the proposed solution? Wouldn't these interventions simply tell us what we already know; i.e., that once the PSP process is initiated, there is a complex causal dance that involves all the putative components of the HoloFoldit system, and that this dance culminates in a fully folded protein molecule that represents the output of the PSP process? If the answer to this question is "yes," then it isn't clear why we need to perform the experimental interventions referred to by the MIE account. Nor is it clear that we need to await the implementation of a real-world HoloFoldit system in order to confirm the constitutive relevance of (e.g.) the HoloLens device. The reason for this is that we already know what the causal connections between the various components of the HoloFoldit system will be. This must be the case, for it is precisely the purpose of a concrete engineering effort to bring these causal connections into existence.

At this point, it is worth noting that there is an important difference between the process of *mechanism discovery* and the process of *mechanism implementation*. The mechanism concept is one that is equally applicable to both scientists and engineers, but this does not mean that scientists and engineers confront the same epistemic challenges when it comes to matters of constitutive relevance. The scientist is primarily concerned with the discovery of mechanisms—they seek to uncover the causal structure of the world using whatever methods they have at their disposal. The engineer, by contrast, is more in the business of mechanism implementation. Their goal is not so much to discover the causal structure of some pre-existing system (although that is sometimes the case); it is more to add causal structures to the world via the construction of particular mechanisms. This distinction is important, for the problem of constitutive relevance, as it is presented in the philosophy of science, is primarily concerned with issues arising as part of the process of mechanism discovery. In this discovery-oriented context, the problem of constitutive relevance is a genuine problem, for scientists often confront naturally-occurring systems and phenomena whose inner workings are unknown to them. For the engineer, however, the "problem" of constitutive relevance looks rather different. The challenge for the engineer is not so much to discover relations of constitutive relevance; it is more to bring these relations into existence. For the engineer (the maker of mechanisms), it makes little sense to rely on experimental interventions as a means of revealing the causal structure of a mechanism, for the causal structure of a mechanism is something that is resolved as part of the design and development effort.

The reason we are encountering this shift in perspective—from mechanism discovery to mechanism implementation—is due to the nature of the HoloFoldit case. In particular, the HoloFoldit case is *not* depicting a naturally occurring cognitive system whose underlying causal structure is utterly unknown to us. It is, instead, depicting a system that might be brought into existence as the result of a deliberate attempt at cognitive systems engineering. Relative to the nature of this engineering effort, the problem of constitutive relevance effectively disappears, for the maker of mechanisms (the engineer) has a degree of insight, understanding, and control that is seldom afforded to the discoverer of mechanisms. Consider, for example, that if the results of experimental interventions were to suggest the absence of a given causal connection, then the engineer could always tweak the design of the HoloFoldit system so as to bring this causal connection into existence. Within certain limits, mechanisms can be tailored to suit the demands of different theoretical accounts of constitutive relevance, so even if the MIE account should not be the final word on constitutive relevance, this does not mean that the "problem" of constitutive relevance will suddenly reappear. In principle, at least, engineers could just heed what philosophers tell them constitutive relevance is, and then proceed to fit the causal profile of a system to match the proposed theoretical constraints. In practice, of course, this is unlikely to happen, but the point, for present purposes, is simply that the problem of constitutive relevance is one that arises in respect of the effort to formulate mechanistic *explanations* of phenomena. In engineering, however, the goal, for the most part, is not to explain a phenomenon; it is more to instantiate a phenomenon—to build a system that makes a phenomenon materially possible. This effort is not without its epistemic issues and challenges, but such issues and challenges are not the same as those that accompany the scientific effort to disclose the causal structure of the world.

In the case of the HoloFoldit system, then, the "problem" of constitutive relevance is "resolved" courtesy of the particular nature of the HoloFoldit system. Given that the HoloFoldit system is a system that is designed to support the PSP process, we do not need to puzzle over the mechanisms that constitute/realize this process. In essence, we already know what the components of the PSP mechanism are because this particular mechanism was specifically designed to realize the phenomenon that interests us in the HoloFoldit case (i.e., the PSP process).[12] Accordingly, I suggest that the HoloLens, holograms, human individual, and the online resources ought to be seen as part of the mechanism that realizes the PSP process. At the very least, I can see no reason why the two component-related criteria discussed in Section 3 should not be satisfied by the HoloFoldit case. That is to say, I can see no reason why the human individual (criterion 2) and the extra-organismic resources (criterion 3) should not be seen to form part of a common, materially hybrid mechanism that serves as the mechanistic realization base for the PSP process.

6 The Problem of Cognitive Ownership

Inasmuch as one accepts the claims made in the previous two sections, then we will have met three of the four criteria presented in Section 3. In particular, we will have satisfied the cognitive status criterion (the PSP process has been identified as a *bona fide* cognitive process); the human component criterion (a human individual is a component of the mechanism responsible for the PSP process); and the extra-organismic component criterion (at least one of the components of the mechanism responsible for the PSP process qualifies as an extra-organismic resource). The remaining criterion is the ownership criterion. As we saw in Section 3, this criterion presents us with an ownership-related issue: Somehow, we have to tie the PSP process to the human individual, such that the process is seen to "belong to" the human individual.[13] What it means for a cognitive process (or some other cognitive phenomenon)

[12] The PSP mechanism is an example of what Glennan and Illari (2018b) dub a mechanism with a designed-and-built etiology. Such mechanisms are to be contrasted with naturally occurring mechanisms whose causal structure does not arise as the result of a deliberate engineering effort. This distinction is important because engineers typically know a great deal about the causal structure of mechanisms that they themselves create. In particular, engineers create mechanisms by selecting components and configuring and/or constraining the causal interactions between these components. Given this, it should be clear that engineers are in a somewhat epistemically privileged position as regards the problem of constitutive relevance. If the components of a mechanism have been selected as part of a design and implementation effort, then there is no need to subject the mechanism to empirical scrutiny as a means of re-discovering these components. Having said this, it is important to note that a mechanism with a designed-and-built etiology is not the same as a mechanism that simply includes one or more engineered components. A pen is an engineered artifact, but it can be used in all manner of mechanisms whose causal structures are unknown to us. What matters here is knowledge about a *mechanism's* causal structure, not the mere fact that this causal structure includes one or more engineered components.

[13] As noted by an anonymous referee, there are multiple ways of understanding the term "cognitive ownership." Carter (2020), for example, refers to a form of cognitive ownership that rests on the "epistemically respectable" endorsement of information emanating from a bio-external resource. This is not the sort of cognitive ownership that concerns us here. In the present paper, we are concerned with whether a putatively extended cognitive process can be seen to belong to a particular cognitive agent.

to belong to a human individual remains a little unclear; nevertheless, a variety of authors have drawn attention to this sort of issue (Roberts, 2012a, 2012b; Rowlands, 2009; Tollefsen, 2006; Wilson, 2004). In order to confirm the status of the PSP process as an extended process (or extended phenomenon), we thus need to tackle the problem of cognitive ownership. In essence, we need to state why we think the PSP process ought to be seen to belong to the human individual (i.e., the user of the HoloFoldit application) despite the fact that the mechanism responsible for the PSP process includes components that lie external to the biological borders of this individual.

There are, I think, a number of ways that we might approach the problem of cognitive ownership. Firstly, we might direct our attention to the causal history of the mechanism underlying the PSP process. In effect, we might ask ourselves "who or what is responsible for this particular mechanism coming into existence?" The answer, in this case, is surely the human individual. This is not to say that the human individual is responsible for the creation of the components that comprise the mechanism (the human individual need not have been responsible for the assembly of the HoloLens device, for example). Rather, the claim is that the human individual plays a causal role in triggering the token instantiation of a mechanism that is responsible (in a constitutive sense) for a token occurrence of the PSP process. It is thus the human individual who (courtesy of their actions) brings the mechanism into existence—they are the ones who switch on the HoloLens device, launch the HoloFoldit application, and ultimately initiate the PSP process. In this sense, the human individual might be seen to "own" the PSP process on account of the fact that they are the ones who are responsible for the creation (or instantiation) of the mechanism that realizes the PSP process.

A second approach to the problem of cognitive ownership is to appeal to the role of the human individual in controlling the overall shape of the PSP process. In this case, we ask ourselves "who or what is responsible for the time-variant causal structure of the mechanism that realizes the PSP process?" In answering this question, our attention is surely drawn to the activities of the human individual. This is not to say that the human individual is unaffected (in a causal sense) by the activities of other components of the same mechanism (e.g., the HoloLens device). Rather, the claim is that the human individual plays a prominent role in shaping the overall activity (or behavior) of the mechanism that realizes the PSP process.

This idea is a little difficult to pin down, since all the components of the PSP mechanism are clearly embedded in a web of causal relations that includes the other components of the mechanism (this, recall, was the main thrust of the argument in Section 5). Nevertheless, I think it is possible to regard the human individual as a prominent locus of control and coordination relative to the time-variant causal structure of the mechanism. This resembles an argument made by Wilson (2004) with respect to the notion of extended memory. Wilson thus suggests that extra-organismic resources:

> […] become integrated into an overall cognitive system that we control, and that control is critical to cognition being ours and to bearing on our lives as agents. Each of us forms a core part of a specific wide memory system, one in

which we serve as a locus of control. And that is why the individual remains the entity that has memories, even if memory is neither taxonomically nor locationally individualistic. (Wilson, 2004, pp. 197–198)

One way of making sense of this appeal to issues of control and coordination is to ask ourselves "whether a human individual ought to be credited with the success or failure of a cognitive routine?" This is the sort of strategy pursued by Roberts (2012b). In this case, we turn our attention to the outcome of the PSP process and ask ourselves whether the human individual ought to be credited with (e.g.) the successful discovery of the protein's native structure. As noted by Roberts, this sort of approach establishes an interesting point of contact with issues of credit attribution in virtue epistemology (see Pritchard, 2010).

Issues of creation, control, and (perhaps) credit thus look to be important in resolving the problem of cognitive ownership. There is no doubt much more that could be said about these issues; for present purposes, however, I want to suggest that none of these issues pose much of a problem for the HoloFoldit case. In respect of creation, for example, it is clearly the human individual that is responsible (in a causal sense) for the creation/instantiation of the mechanism that is then responsible (in a constitutive sense) for the PSP process. The human individual also plays a role in shaping (and thus controlling) the flow of information within this world-involving cognitive circuit. Finally, if a successful solution to the protein folding problem were to be discovered, I suspect we would have little problem in crediting the human individual with this discovery. In all likelihood, we would recognize the discovery of the protein's native conformation as a form of *cognitive* achievement, and we would, I think, see this achievement as being due, at least in part, to the cognitive and epistemic wherewithal of the biologically-bounded human individual. Personally, then, I can see no reason to demur from the idea that the PSP process ought to be seen to "belong to" the human individual.[14]

[14] As noted in Section 3, the present paper is concerned with a particular form of extended cognition, glossed as human-centered extended cognition. There is, however, another form of cognitive extension that is potentially relevant to the HoloFoldit case. This is what has been dubbed human-extended machine cognition (Smart, 2018). In essence, human-extended machine cognition is a form of cognitive extension that is centered on an intelligent, technological system, such as an AI system. In the HoloFoldit case, this raises the possibility that the HoloLens device might be seen to own the PSP process. That is to say, once we allow for the idea that a machine (or technological artifact) could be the subject of cognitive extension, then it is perfectly reasonable to ask why the human ought to be seen to incorporate the HoloLens into *their* routines, as opposed to the HoloLens incorporating the human individual into *its* routines. In response to this issue, it is worth noting the way in which the overall trajectory of the PSP process—understood as a succession of intermediate problem states—is governed by the thoughts and actions of the human individual. It is thus the human individual that effects the transition from one problem state to another by, in effect, choosing what actions to perform, what constraints to set, what solution candidates to explore, and so on. The HoloLens, by contrast, isn't really involved in governing the trajectory of the problem-solving process. To be sure, the HoloLens responds to what the human does, but it doesn't "decide" whether to respond this way or that. Nor does the HoloFoldit app recommend that the user undertake specific actions in response to the current problem state. If this were to be the case—if, for example, the HoloFoldit app, were to start guiding the human user through the problem-solving process by recommending specific actions—then I suspect our ownership-related intuitions might begin to shift. (I am grateful to an anonymous referee for bringing this issue to my attention.)

7 Discussion

The philosophical and cognitive scientific literature provides us with a number of putative cases of extended cognizing. For the most part, however, these cases involve the use of technologically low-grade resources, such as the use of pen and paper resources to solve long multiplication problems (e.g., Wheeler, 2010). The aim of the present paper was to describe a novel case of extended cognizing based around a technologically-advanced resource, namely, the Microsoft HoloLens. This particular case—the HoloFoldit case—demonstrates how active externalist theorizing might be applied to situations involving mixed, virtual, and augmented reality technologies. In this sense, the HoloFoldit case speaks to the recent philosophical interest in virtual/mixed reality (e.g., Chalmers, 2022; Lanier, 2017; Metzinger, 2018; Smart, 2021; Turner, 2022), as well as the emerging philosophical/technological interest in what has been dubbed the Metaverse (e.g., Mystakidis, 2022).

More generally, the HoloFoldit case shows us how the issues and concerns of active externalists might dovetail with those working in a number of other disciplinary areas. The status of holograms as cinematic (or "moving" image) resources, for example, highlights a point of interdisciplinary contact with those working in cinematic philosophy (e.g., Shamir, 2016), while the mechanistic orientation of the present paper highlights how the theoretical resources of neo-mechanical philosophy (e.g., Glennan & Illari, 2018a) might be used to evaluate putative cases of extended cognizing. Finally, it should be clear that the HoloFoldit case is of considerable epistemic importance. By solving the protein folding problem, we improve our (mechanistic) knowledge and understanding of a rich array of biological processes (both normal and abnormal)—protein molecules are, of course, the lifeblood of practically all biological processes. In this sense, the HoloFoldit case is well-placed to pique the interests of those who are concerned with the epistemological significance of active externalist claims (e.g., Carter et al., 2018).

While it is possible to regard the HoloFoldit case as nothing more than a technologically advanced form of extended cognizing, there are, I think, a number of features that make the case of particular philosophical interest. Note, for example, that the HoloFoldit case involves the use of online (i.e., Internet-accessible) computational routines. Inasmuch as we regard these computational routines as part of the PSP process (i.e., constituents of the mechanism that is responsible for the PSP process), then the HoloFoldit case provides us with an example of Internet-extended cognition, i.e., a form of extended cognition that involves the incorporation of online resources into cognitive routines (see Smart, 2017).[15]

Another issue raised by the HoloFoldit case concerns the virtual status of the holographic resources that are manipulated as part of the protein folding task. In

[15] Note the way in which online resources are being incorporated into a cognitive routine is markedly different from that associated with more conventional forms of Internet-extended cognition (e.g., situations in which a human individual accesses online content via a Web browser interface). In the HoloFoldit case, the human user may be unaware that online services are being invoked as part of the problem-solving process. This is unlike the more conventional cases, where the user is typically aware that they are accessing online content.

particular, the HoloFoldit case introduces us to the idea that virtual objects (in this case, holograms) might form part of the mechanisms that realize cognitive processes. In essence, the HoloFoldit case introduces us to the idea of hologrammatically extended cognition, i.e., a form of cognitive extension that involves the incorporation of holographic (or, more generally, virtual) resources into human cognitive routines (see Smart, 2020). The extent to which such claims should be seen as controversial remains unclear, and much may depend on how we conceptualize the notion of a virtual object (see Chalmers, 2017, for more on this).[16] Nevertheless, issues of cognitive extension are typically discussed with reference to concrete, physical resources, such as notebooks and iPhones (e.g., Clark, 2008). The primary concern in such cases is whether or not we should talk of some physical extra-organismic resource as forming part of the mechanistically relevant fabric that realizes some cognitive routine. This issue remains important in the HoloFoldit case; but we are now confronted with an additional concern: to what extent does it make sense to talk of a virtual, cinematic resource (e.g., a hologram) as forming part of the *physical* machinery of the mind? Does this mean that hologrammatically extended cognitive routines are apt to emerge from the forms of photic flux that define the moving image, and what does this mean for our current understanding of the cognitive and philosophical significance of the cinematic medium?

A further peculiarity of the HoloFoldit case is that it directs our attention to issues that are not so readily apparent in the more familiar cases of cognitive extension; i.e., cases that limit their attention to the realm of physical objects (and physical reality). Consider, by way of an example, the claim that mere perceptual objects ought not to be included in an extended cognitive mechanism on the grounds that there is no mutual interaction (or reciprocal causal influence) between the perceptual object and the human observer (Palermos, 2014). Palermos (2014, p. 31) cites the example of a human observing a tree. As the human moves around the tree, the human's perspective of the tree changes, but the tree itself is not affected by the observer's movements. Accordingly, the tree (*qua* perceptual object) cannot be part of any of the cognitive routines associated with the human individual because there is no evidence of *reciprocal* causal influence between the observer and tree—the user might be affected by the tree, but the tree is not subject to the causal influences exerted by the observer's movements.

[16] In the HoloFoldit case, it is not unreasonable to regard holograms as a form of photonic object—an "object" comprised of light or photons. Such a conceptualization challenges the idea that virtual objects ought to be seen as digital objects, which is the position endorsed by Chalmers (2017). As a photonic object, it seems appropriate to regard holograms as possessing certain causal powers, which is a central tenet of virtual realism. But we can make sense of this idea without identifying holograms as digital objects. McDonnell and Wildman (2019) dispute the claim that virtual objects have genuine causal powers, but their analysis is mostly geared to causal interactions *between* virtual objects as opposed to the causal influence exerted by virtual objects on conventional physical objects. There is, to be sure, much more that could be said about the ontic character of holograms. For present purposes, however, what matters is that a consideration of the HoloFoldit case introduces us to the possibility that holograms may count as the causally active constituent of extended cognitive circuits, even if such entities are unlike those that typically feature as part of mechanistically oriented approaches to the study of cognitive scientific phenomena.

Palermos (2014) is, I think, quite correct to doubt the constitutive relevance of a mere perceptual object when that object qualifies as a conventional physical object. Such claims seem much more dubious, however, when it comes to the realm of virtual objects. In the HoloFoldit case, the object of perception is the holographic protein molecule, and this object needs to be continuously updated based on where the observer is and what they are doing. In this case, the act of walking around the holographic protein molecule *does* involve a causal link between the human observer and the perceptual object, albeit one that is mediated by the HoloLens device. Given this, it seems, that perceptual objects might, under at least some circumstances, qualify as part of an extended cognitive routine.

In addition to the issues raised by the use of a mixed reality device, the present paper shows how a mechanistic approach to extended cognition might be used to evaluate putative cases of extended cognition. In part, this approach was inspired by the interest in mechanistic concepts within the active externalist literature (e.g., Fazekas, 2013; Kaplan, 2012; Smart, in press). The approach was also inspired by the *synthetic* nature of the HoloFoldit case—the fact that we are being asked to imagine a system that might be designed and developed so as to support a particular form of extended cognizing. This is important, for it encourages us to adopt a somewhat different approach to mechanisms than that seen in the bulk of the mechanistically oriented philosophical literature. In particular, rather than see mechanisms as something that need to be discovered and described as part of the process of *mechanism discovery*, the HoloFoldit case presents us with a system whose mechanisms have already been "discovered" and described as part of the process of *mechanism implementation*. This potentially alters the way we think about the problem of constitutive relevance, which is a perennial feature of debates and discussions in both the philosophy of mind and the philosophy of science. In particular, if we already know how a cognitive process is realized because we, ourselves, build the mechanism that sustains the process, then it is not particularly clear that we still confront the problem of constitutive relevance. In such cases, I suggest, we already know what the components of the relevant mechanism are, so there is no need to worry about the ostensible shortcomings of interventionist-style manipulations to evaluate putative cases of cognitive extension (see Baumgartner & Wilutzky, 2017).

There is, no doubt, much more to be said about this particular issue (as well as other issues raised throughout the paper). For present purposes, however, I hope to have shown that a consideration of the HoloFoldit case provides us with a novel target for philosophical debates about extended cognition, especially as those debates relate to theories of mechanistic explanation and mechanistic realization/constitution. By tackling the issues raised by the HoloFoldit case, we may come to a better understanding of the way in which emerging technologies are poised to liberate the human mind from its cranial confines, extending its reach to the wider physical world in which we currently live, as well as the emerging virtual worlds that we are just beginning to inhabit.

Acknowledgements The author would like to thank three anonymous reviewers for their helpful and insightful remarks on an earlier version of the manuscript.

Author Contribution The author (PS) assumed sole responsibility for all aspects of the work presented herein.

Funding This work is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) as part of the PETRAS National Centre of Excellence for IoT Systems Cybersecurity under Grant Number EP/S035362/1.

Data availability Not applicable.

Code Availability Not applicable.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication The author consents to the publication of this article.

Competing Interests The author declares no competing interests.

References

Adams, F. (2010). Why we still need a mark of the cognitive. *Cognitive Systems Research*, *11*(4), 324–331.

Adams, F., & Aizawa, K. (2008). *The bounds of cognition*. Blackwell.

Adams, F., & Garrison, R. (2013). The mark of the cognitive. *Minds and Machines*, *23*(3), 339–352.

Baumgartner, M., & Casini, L. (2017). An abductive theory of constitution. *Philosophy of Science*, *84*(2), 214–233.

Baumgartner, M., & Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *The British Journal for the Philosophy of Science*, *67*(3), 731–756.

Baumgartner, M., & Wilutzky, W. (2017). Is it possible to experimentally determine the extension of cognition? *Philosophical Psychology*, *30*(8), 1104–1125.

Baumgartner, M., Casini, L., & Krickel, B. (2020). Horizontal surgicality and mechanistic constitution. *Erkenntnis*, *85*, 417–340.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C*, *36*(2), 421–441.

Carter, A. J., Clark, A., Kallestrup, J., Palermos, O. S., & Pritchard, D. (Eds.) (2018). *Extended epistemology*. Oxford University Press.

Carter, J. A. (2020). Intellectual autonomy, epistemic dependence and cognitive enhancement. *Synthese*, *197*(7), 2937–2961.

Cassidy, K. C., Šefčík, J., Raghav, Y., Chang, A., & Durrant, J. D. (2020). ProteinVR: Web-based molecular visualization in virtual reality. *PLoS Computational Biology*, *16*(3), e1007747.

Chalmers, D. J. (2017). The virtual and the real. *Disputatio*, *9*(46), 309–352.

Chalmers, D. J. (2022). *Reality+: Virtual worlds and the problems of philosophy*. W. W. Norton & Company.
Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press.
Clark, A. (2011). Finding the mind. *Philosophical Studies*, *152* (3), 447–461.
Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19.
Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popović, Z., & Foldit Players (2010). Predicting protein structures with a multiplayer online game. *Nature*, *466*(7307), 756–760.
Craver, C. (2007a). Constitutive explanatory relevance. *Journal of Philosophical Research*, *32*, 3–20.
Craver, C. (2007b). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Clarendon Press.
Craver, C., & Tabery, J. (2016). Mechanisms in science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.* Stanford University.
Craver, C. F., Glennan, S., & Povich, M. (2021). Constitutive relevance & mutual manipulability revisited. *Synthese*, *199*(3), 8807–8828.
Dill, K. A., & MacCallum, J. L. (2012). The protein-folding problem, 50 years on. *Science*, *338*(6110), 1042–1046.
Fazekas, P. (2013). The extended mind thesis and mechanistic explanations. In D. Moyal-Sharrock, V. A. Munz, & A. Coliva (Eds.), *Mind, Language, and Action* (pp. 125–127). Austrian Ludwig Wittgenstein Society.
Fu, W. T., Hills, T., & Todd, P. M. (2015). Interfacing mind and environment: The central role of search in cognition. *Topics in Cognitive Science*, *7* (3), 384–390.
Glennan, S. (2017). *The new mechanical philosophy*. Oxford University Press.
Glennan, S., & Illari, P. M. (Eds.) (2018a). *The Routledge handbook of mechanisms and mechanical philosophy*. Routledge.
Glennan, S., & Illari, P. M. (2018b). Varieties of mechanisms. In S. Glennan & P. M. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 91–103). Routledge.
Goldstone, R. L., & Theiner, G. (2017). The multiple, interacting levels of cognitive systems (MILCS) perspective on group cognition. *Philosophical Psychology*, *30*(3), 338–372.
Harinen, T. (2018). Mutual manipulability and causal inbetweenness. *Synthese*, *195*(1), 35–54.
Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology*, *36*, 49–67.
Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., & CSR Group (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences*, *19*(1), 46–54.
Hoffman, M. A., & Provance, J. B. (2017). Visualization of molecular structures using HoloLens-based augmented reality. *AMIA Joint Summits on Translational Science*. American Medical Informatics Association.
Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, *2*(1), 119–135.
Japyassú, H. F., & Laland, K. N. (2017). Extended spider cognition. *Animal Cognition*, *20*(3), 375–395.
Kaiser, M. I., & Krickel, B. (2017). The metaphysics of constitutive mechanistic phenomena. *The British Journal for the Philosophy of Science*, *68*(3), 745–779.
Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biology & Philosophy*, *27*(4), 545–570.
Kersten, L. (2022). A new mark of the cognitive? Predictive processing and extended cognition. *Synthese*, *200*(4), 1–25.
Khatib, F., Cooper, S., Tyka, M. D., Xu, K., Makedon, I., Popović, Z., Baker, D., & Foldit Players (2011). Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences*, *108*(47), 18949–18953.
Kirsh, D. (2009). Problem solving and situated cognition. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 264–306). Cambridge University Press.
Kiverstein, J., & Sims, M. (2021). Is free-energy minimisation the mark of the cognitive? *Biology & Philosophy*, *36*(Article 25), 1–27.
Klau, G. W., Lesh, N., Marks, J., & Mitzenmacher, M. (2010). Human-guided search. *Journal of Heuristics*, *16*(3), 289–310.
Krickel, B. (2017). Making sense of interlevel causation in mechanisms from a metaphysical perspective. *Journal for General Philosophy of Science*, *48*(3), 453–468.
Lanier, J. (2017). *Dawn of the new everything: A journey through virtual reality*. Henry Holt and Company.
McDonnell, N., & Wildman, N. (2019). Virtual reality: Digital or fictional? *Disputatio*, *11*(55), 371–397.

Metzinger, T. K. (2018). Why is virtual reality interesting for philosophers? *Frontiers in Robotics and AI*, *5*(Article 101), 1–19.

Miłkowski, M., Clowes, R., Rucińska, Z., Przegalińska, A., Zawidzki, T., Krueger, J., Gies, A., McGann, M., Afeltowicz, L., Wachowski, W., Stjernberg, F., Loughlin, V., & Hohol, M. (2018). From wide cognition to mechanisms. *Frontiers in Psychology*, *9*(Article 2393), 1–17.

Müller, C., Krone, M., Huber, M., Biener, V., Herr, D., Koch, S., Reina, G., Weiskopf, D., & Ertl, T. (2018). Interactive molecular graphics for augmented reality using HoloLens. *Journal of Integrative Bioinformatics*, *15*(2), 1–13.

Mystakidis, S. (2022). Metaverse. *Encyclopedia*, *2*(1), 486–497.

Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.

Newell, A., & Simon, H. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the Association of Computing Machinery*, *19*(3), 113–126.

Palermos, S. O. (2014). Loops, constitution, and cognitive extension. *Cognitive Systems Research*, *27*, 25–41.

Parise, A. G., Gagliano, M., & Souza, G. M. (2020). Extended cognition in plants: Is it possible? *Plant Signaling & Behavior*, *15*(2), 1710661.

Povich, M. (in press). Mechanistic explanation in psychology. In H. J. Stam & H. L. de Jong (Eds.), *The SAGE handbook of theoretical psychology*. SAGE Publications Ltd.

Pritchard, D. (2010). Cognitive ability and the extended cognition thesis. *Synthese*, *175*(1), 133–151.

Prychitko, E. (2021). The causal situationist account of constitutive relevance. *Synthese*, *198*, 1829–1843.

Roberts, T. (2012a). Taking responsibility for cognitive extension. *Philosophical Psychology*, *25*(4), 491–501.

Roberts, T. (2012b). You do the maths: Rules, extension, and cognitive responsibility. *Philosophical Explorations*, *15*(2), 133–145.

Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philosophical Psychology*, *22*(1), 1–19.

Shamir, T. S. (2016). *Cinematic philosophy*. Palgrave Macmillan.

Simon, H. A. (1996). *The sciences of the artificial*, 3rd edn. MIT Press.

Smart, P. R. (2017). Extended cognition and the internet: A review of current issues and controversies. *Philosophy & Technology*, *30*(3), 357–390.

Smart, P. R. (2018). Human-extended machine cognition. *Cognitive Systems Research*, *49*, 9–23.

Smart, P. R. (2020). The Joi of holograms. In T. Shanahan & P. R. Smart (Eds.), *Blade runner 2049: A philosophical exploration* (pp. 127–148). Routledge.

Smart, P. R. (2021). Shedding light on the extended mind: HoloLens, holograms, and internet-extended knowledge. *Frontiers in Psychology*, *12*, 1–16.

Smart, P. R. (in press). Toward a mechanistic account of extended cognition. *Philosophical Psychology*.

Tollefsen, D. P. (2006). From extended mind to collective mind. *Cognitive Systems Research*, *7*(2–3), 140–150.

Turner, C. (2022). Augmented reality, augmented epistemology, and the real-world web. *Philosophy & Technology*, *35*, 1–28.

van Eck, D. (2019). Constitutive relevance in cognitive science: The case of eye movements and cognitive mechanisms. *Studies in History and Philosophy of Science Part A*, *73*, 44–53.

van Eck, D., & de Jong, H. L. (2016). Mechanistic explanation, cognitive systems demarcation, and extended cognition. *Studies in History and Philosophy of Science Part A*, *59*, 11–21.

Wheeler, M. (2010). In defense of extended functionalism. In R. Menary (Ed.), *The extended mind* (pp. 245–270). MIT Press.

Wheeler, M. (2019). Breaking the waves: Beyond parity and complementarity in the arguments for extended cognition. In M. Colombo, E. Irvine, & M. Stapleton (Eds.), *Andy Clark and His Critics* (pp. 81–95). Oxford University Press.

Wilson, R. A. (2004). *Boundaries of the mind: The individual in the fragile sciences: Cognition*. Cambridge University Press.

Wilson, R. A., & Craver, C. (2007). Realization: Metaphysical and scientific perspectives. In P. Thagard (Ed.), *Philosophy of psychology and cognitive science* (pp. 81–104). North-Holland.

Zednik, C. (2011). The nature of dynamical explanation. *Philosophy of Science, 78*(2), 238–263.