European Journal for Philosophy of Science (2023) 13:41
https://doi.org/10.1007/s13194-023-00547-4

PAPER IN THE PHILOSOPHY OF THE SCIENCES OF MIND AND BRAIN

Minimal model explanations of cognition

Nick Brancazio[1] · **Russell Meyer**[2]

Received: 5 October 2022 / Accepted: 14 August 2023 / Published online: 29 August 2023

Abstract

Active materials are self-propelled non-living entities which, in some circumstances, exhibit a number of cognitively interesting behaviors such as gradient-following, avoiding obstacles, signaling and group coordination. This has led to scientific and philosophical discussion of whether this may make them useful as minimal models of cognition (Hanczyc, 2014; McGivern, 2019). Batterman and Rice (2014) have argued that what makes a minimal model explanatory is that the model is ultimately in the same universality class as the target system, which underpins why it exhibits the same macrobehavior. We appeal to recent research in basal cognition (Lyon et al., 2021) to establish appropriate target systems and essential features of cognition as a target of modeling. Looking at self-propelled oil droplets, a type of active material, we do not find that organization alone indicates that these systems exhibit the essential features of cognition. We then examine the specific behaviors of oil droplets but also fail to find that these demonstrate the essential features of cognition. Without a universality class, Batterman & Rice's account of the explanatory power of minimal models simply does not apply to cognition. However, we also want to stress that it is not intended to; cognition is not the same type of behavioral phenomena as those found in physics. We then look to the minimal cognition methodology of Beer (1996, 2020a, b) to show how active materials can be explanatorily valuable regardless of their cognitive status because they engage in specific behaviors that have traditionally been expected to involve internal representational dynamics, revealing misconceptions about the cognitive underpinnings of certain, specific behaviors in target systems where such behaviors are cognitive. Further, Beer's models can also be genuinely explanatory by providing dynamical explanations.

Keywords Active materials · Minimal cognition · Minimal models · Universality class · Cognitive processes

---

Extended author information available on the last page of the article

1 Introduction

Active materials are self-propelled synthetic entities which, in some circumstances, exhibit a number of interesting behaviors such as gradient-following, avoiding obstacles, signaling and group coordination (Hagan & Baskaran, 2016; Needleman & Dogic, 2017; Horibe et al., 2011). What sets active materials apart from other synthetic materials is that they are created and studied primarily "for the targeted manipulation of [their] active behaviors" (Bursten, 2020, p. 2011-12). The behaviors of these materials have been generating increasing scientific and philosophical interest because of their similarity to behaviors thought to indicate cognition in basal organisms. For instance, most readers might be familiar with chemotactic behavior in bacteria, but be unaware that analog chemotactic behaviors can be generated in some kinds of engineered oil droplets (Hanczyc, 2011). Without prior knowledge that these are not living systems, we might think that they are undertaking the same kinds of activities, such as chasing sustenance or avoiding harm.

Because of the resemblances between some of active materials' behaviors and those we find in living systems, it has been suggested that these active materials can provide useful models of cognition in more complex systems (Bich & Moreno, 2016; McGivern, 2019; Hanczyc & Ikegami, 2010). One might see these as fairly innocuous general claims. As Godfrey-Smith (2006) would call it, *the strategy of model-based science* is to use models in order to build theories–and as active materials research is still in a very early stage, this seems highly appropriate. Godfrey-Smith's starting point is Giere's (1988) account of modeling, in which models are built for understanding the world through resemblance relations, where "[t]he modeler's strategy is to gain understanding of a complex real-world system via an understanding of a simpler, hypothetical system that resembles it in relevant respects" (Godfrey-Smith, 2006, p. 726). Perhaps the idea is just that active materials give us some increased understanding of higher level cognition in more complex systems that can be useful for building theories.

However, why and whether it is that we are able to gain understanding in this way has been brought under scrutiny by philosophers of science. Batterman and Rice (2014) develop an account of minimal model explanations, which explain by producing the same macroscale behaviors as the target phenomenon by isolating the features that facilitate the behavior. To have explanatory power, then, the model need not accurately "represent" all of the microscale phenomena; it only needs to generate the same kind of macrobehavior by employing a minimal number of mechanisms or processes. This stands in contrast with accounts of the explanatory power of mechanistic models, in which explanation involves a strict accuracy condition between the model's causal mechanism and the causal mechanism of the target phenomenon. Mechanistic models aim to describe the causal relations between components that make up a mechanism-in-the-world, and then show how said mechanism gives rise to the phenomenon of interest (Bechtel & Abrahamsen, 2006, 2011; Craver, 2007). On Batterman and Rice's account, a minimal model is explanatory through replication of the macrobehavior even though the microdetails, such as specific mechanisms, may differ.

The theory of minimal models developed by Batterman and Rice (2014) would suggest that for active materials to be minimal models of cognition, they would need to exhibit the *essential features* of the cognitive behaviors of a target system (McGivern, 2019). Batterman and Rice claim that in "common features accounts" of the explanatory power of minimal models, it is a common causal, topological, mathematical, or mechanistic feature that makes the model able to accurately explain the target phenomenon. They argue in response that accurately capturing a specific feature of the phenomenon is not at the basis of its explanatory power; instead, it is in virtue of the macrobehavior of both the target system and model system belonging in the same universality class. Minimal models are not concerned with isolating underlying causes under some criteria of accuracy and completeness; rather, we can identify the commonalities associated with a macroscale behavior in the target system without needing to involve many of the microdetails of the target system. Looking at models used in physics, Batterman and Rice (2014) argue that the reason these minimal models are explanatorily successful is because the models themselves are a member of the same universality class (Goldenfeld, 1992/2018) as the phenomena they seek to model.[1] The universality class is the set that exhibits the phenomenon regardless of the physical details of the instantiation of that phenomenon. The essential features of the behavior belong to all systems in that same universality class–and the behavior found in the model itself shares these essential features rather than simply representing them. Using minimal models of fluid flow as an example (2014), they show that these models include a number of lattice nodes, far less than the actual number of particles that might be flowing in fluid, in order to demonstrate how fluid will flow in different environmental conditions. Both the model system and the fluid itself have the "essential features" of flow (locality, conservation, symmetry) needed in order to qualify as members of the universality class, though the model contains a minimal number of nodes (rather than all of the involved particles) to instantiate the behavior. After all, a minimal model captures the essential features of a behavior in the most economical fashion (Goldenfeld, 1992/2018). The minimal model is able to exhibit the macrobehavior precisely because it has the same essential features, thus making the model *itself* an instance of the type of thing that exhibits the behavior.

In evaluating claims that active materials might be thought of as minimal models of cognition, we think some clarification is needed on the essential features of cognition and their relation to the kinds of macrobehaviors in which active materials can engage. An oil droplet engaging in chemotactic behavior, for example, might be thought to be a minimal model of bacterial chemotaxis because it engages in the behavior though it has only a few chemical components (Hanczyc, 2014; Hanczyc & Ikegami, 2010). If this is a cognitive behavior in bacteria, and an active material can perform the same behavior, then we might be tempted to treat these as belonging in the same universality class of cognitive behaviors. Given this, McGivern (2019)

---

[1] Explanatory success, for minimal models of this type, is achieved when the model captures the universal behavior accurately or correctly (with a caveat that these are philosophically debatable terms). The minimal model, unlike mechanistic models, isolates the essential features of the phenomena in a way that gives purchase on the phenomena at other scales. For example, Batterman and Rice (2014) refer to the explanatory success of the Navier-Stokes equation, which scales up accurately enough to guide work in engineering, aeronautics, and so forth. We thank an anonymous reviewer for urging us to clarify this.

argues that “a proper understanding of minimal models forces us to embrace the idea that if these systems are informative as models of cognition, they should also count as instances of cognition” (p. 444). Here, we disagree with McGivern, but we take up the challenge he proposes. If one subscribes to the Batterman and Rice view of how minimal models are explanatorily justified, then we have to contend with a dilemma: “either these non-living systems are valuable as minimal models [of cognition] or they are not. If they are valuable as minimal models, then they must be instances of cognition; if they are not instances of cognition, then they are not valuable minimal models of cognition either” (McGivern, 2019, p. 10). The challenge for those that consider active materials to provide useful minimal models of cognition is that either we must have a principle of exclusion that specifies why these minimal models are useful despite not being cognitive, or bite the bullet and take active materials to be instances of cognition.

We argue that it is still an open question how we should understand the resemblance between these systems and living systems, but that it is not one that relies on both the model and target system being cognitive. The minimal models account proposed by Batterman and Rice (2014) does not tell us why minimal models of cognition are explanatorily valuable. We point instead to the minimal modeling methodology of Beer (1996) to show how minimal models of cognition can offer explanatory purchase on cognition without belonging in the same universality class as cognitive systems.

Even if we take claims about the possibility that they can act as minimal models of cognition to be saying that the creation and manipulation of behaviors in active materials can offer us some understanding of the behaviors of cognitive systems, we still require some clarification on the specific relationship between these *behaviors* and cognition itself. In the following section, we appeal to the biological function of cognition as taken up in *basal cognition* (Lyon et al., 2021) to establish appropriate target systems and essential features. In section three, we provide an analysis of the organization of self-propelled oil droplets, but do not find that organization alone indicates that the systems exhibit the essential features of cognition. Section four then evaluates whether instead we can look at the behaviors of motile oil drops as providing minimal models of specific cognitive capacities, but we fail to find that these behaviors demonstrate the essential features of cognition. Because cognitive behaviors are not a universality class, Batterman & Rice’s account of the explanatory power of minimal models simply does not apply. However, we also want to stress that it is not intended to; cognition is not the same type of behavioral phenomena as fluid flow. We then look in section five to the methodology of Beer (1996, 2020a, b) to show how active materials can be minimally cognitively interesting regardless of their cognitive status because they engage in specific behaviors that have traditionally been expected to involve internal representational dynamics. It is not because these model systems are also cognitive that they can be explanatorily valuable, but that they can reveal misconceptions about the cognitive underpinnings of certain, specific behaviors in target systems where such behaviors *are* cognitive. However, Beer’s methodology can also be genuinely explanatory by providing dynamical explanations.

2 Does cognition have essential features?

In order to assess whether active materials can be explanatory as minimal models of cognition, we first need a clear understanding of the essential features of cognition. Defining cognition is, to put it mildly, tricky. On the one hand, as Pamela Lyon explains it, cognition has no agreed upon definition precisely because there is no agreement on its function (2015). Lyon points out that unlike a definition for respiration, which we can articulate from its biological function and which is precisely what enables us to do careful research on diverse implementations of respiration, definitions of cognition seem to flexibly adapt to researchers' goals, theoretical commitments, and intuitions, which have changed over time (see Akagi, 2018). If cognitive science hasn't narrowed in as a community on the function of cognition, how are we to know which behaviors are indicative of cognition, let alone determine cognition's essential features?

On the other hand, there is a long history of anthropocentric argumentation for reserving 'cognition' to a strict subset of folk psychological phenomena for which representational explanations are said to be required (for example, see Adams, 2018; Fulda, 2017; Adams & Aizawa, 2008). These theorists leverage their commitments to specific cognitive frameworks to establish that there are common underlying features of cognition. However, the adoption of a framework previous to establishing what makes a *behavior* cognitive puts the explanatory cart before the horse (Lyon, 2015; Sims, 2021), as it demarcates a subset of possible explanans before clarifying the explanandum. As Colin Allen phrases it, settling disputes over these kinds of demarcation criteria "using using philosophical precepts, such as 'intrinsic intentionality', or (quasi) theoretical terms such as 'representations'… is likely, in my view, to be rendered otiose by scientific developments" (2017, p. 4237). If we instead look at the sciences and "the structural and functional diversity of adaptive information-processing systems" (Allen, 2017, p. 4243), it calls into question the usefulness of arguing over demarcation criteria when "so much remains to be learned about the variety and complexity of cognitive systems" (ibid.).

Rather than defining cognition by pointing to specific underpinnings or operations, the research program of *basal cognition* proposes we study cognition using biological principles, with the understanding that cognition, like other biological processes, serves a functional role aimed at meeting the existential needs of the organism (Lyon et al., 2021; Levin et al., 2021; Abramson & Levin, 2021). To be clear, this is an *approach* to studying cognition, and includes as its advocates a growing number of multidisciplinary empirical scientists studying cognition through systems traditionally out of the purview of mainstream cognitive science, such as plants and bacteria. One benefit to the basal approach to cognition is that it enables us to investigate the workings of cognition as involving integrated biological operations performed for the system in a number of ways, giving researchers an agreed-upon explanandum without a commitment to a single cognitive framework as necessary for providing explanantia. That is, researchers can discuss a wide variety of underpinnings (e.g. mechanisms, processes, sensorimotor contingencies, networks, and/or circuits) across the biological spectrum as enabling cognitive operations, giving a basis for comparison and pluralism rather than preemptively championing any particular one of these components as *the* single source of cognition.[2] Whether one agrees or disagrees with how it carves out the domain of interest, this provides a fruitful way to frame further agreements and disagreements. In other words, framework-specific debates that begin by using this approach to establish their explanatory aim will have an enriched starting point.

It might seem as though basal cognition is being used to establish a cognitive baseline, or "minimal cognition" as such a baseline is often called. The term "minimal cognition" is often used to refer to the minimal instantiation dynamics, capacities, or mechanisms behind cognition in certain systems (e.g. Van Duijn et al., 2006; Bich and Moreno, 2016; see Brancazio et al., 2020 for review). Basal cognition instead proposes that we look at cognition in terms of what it achieves for biological systems without setting any particular instantiation criteria. That is, it is the basis for a research program, not a position on what cognition must be in metaphysical terms. Though there is some literature at the overlap of basal and "minimal" cognition, the adjective of "minimal" will be aside here in relation to cognition for a few reasons. First, we don't want to imply that minimal cognition picks out a particular category, or that it exists in relation to some other "fully-fledged" category (Lyon, 2020). Second, we don't want to confuse the term "minimal cognition" with the shorthand used for Randall Beer's modeling methodology (1996), which we discuss in a further section. Third, we aren't sure how to establish that a living system achieves cognitive behaviors in "minimal" ways. For example, the receptor and signal transduction systems in bacteria are quite complex, especially given their size. A bacterium's chemical detection network contains an estimated 8,000 receptors, made up of a combination of five different types of transmembrane proteins (Jung et al., 2018). Their signal transduction system contains two feedback cycles operating at different timescales, which are used to compare levels of chemical attractants and repellents at the receptors (Blair, 1995). We might be able to say that a *minimal model* achieves a behavior in a simple way, but we are not fully convinced that any living system undertaking a behavior does.

The reader with higher-level framework commitments may yet think that what determines that an organism or system is cognitive is some kind of substrate or content dependance, which we've clarified is a framework commitment, not an approach.[3] We take the onus to be on theorists committed to such high-level criteria to demonstrate that their framework can carve off a unique, non-overlapping subset

---

[2] Basal cognition is open to a host of non-neuronal implementations of cognitive processes. This does not imply computational multiple realizability in the traditional sense, as basal cognition is not committed to any specific framework (computational, FEP, organizational, etc.). Therefore whether or not something such as a representational vehicle is multiply realizable is of no concern to basal cognition as an approach, though it may be of concern to a particular researcher in theorizing about a (basal) sample system through the lens of a specific framework (e.g. computational functionalism). We thank an anonymous reviewer for pushing us to clarify this.

[3] An anonymous reviewer proposes that Colaço's (2022) suggestion that the principles of basal cognition contain claims about what cognition is, treated as conjectures, offers further clarity on the difference between a framework and an approach. We disagree that basal cognition is treated as a conjecture or hypothesis by its proponents in the way these are described by Colaço, and that this misunderstanding stems from Colaço not treating basal cognition as a functional approach capable of accommodating multiple frameworks (in other words, it does not distinguish between proposals for determining explanandum

of processes serving a functional role in order to justify a label of "cognition" distinct from (and more deserving) than these other phenomena.[4] As we use it here, cognition is taken to be a theoretical term that can adapt as needed for use in the sciences (Akagi, 2018). If we instead make the priority to argue that cognitive behaviors are only those that involve certain kinds of underpinnings, and we treat the investigation of behaviors as looking for those specific underpinnings, then we ought to be concerned that any empirical analysis of those behaviors might be an exercise in reinforcement of folk psychological intuitions rather than explanation (Buzsáki, 2019).

In cases where active materials are proposed as possible minimal models of cognition, cognition is generally being used in an inclusive sense that applies to fundamental biological processes across a wide (if not the complete) range of biological individuals. Basal cognition researchers propose that we ought to think of cognition as including processes such as "sensing, information processing, memory, valence, decision making, learning, anticipation, problem solving, generalization and goal directedness" (Lyon et al., 2021, p.1; see also Levin et al., 2021) which serve biological functions and can be diversely and non-neuronally implemented. For those approaching explanations of cognition from this angle, chemotaxis serves as a paradigmatic example of a behavior serving a biologically-based functional role for the organism. In brief, chemotaxis is directed movement up toward a gradient of chemical chemoattractants or away from chemorepellants (Webre et al., 2003). This seemingly simple behavior is said to involve a suite of capacities for bacteria: chemical sensing, a memory mechanism (for comparing current and past chemical saturation in the environment), and the ability to switch between kinds of motor activity (swimming and tumbling). Importantly, chemotactic behavior involves sensorimotor coordination, which has been argued to be the primary domain of cognition (Van Duijn et al., 2006). This is well supported within the sciences: "If a brain is an organ that uses sensory information to control motor activity, then the bacterial nanobrain would fit the definition" (Webre et al., 2003). This inclusion ensures that mere motion can unproblematically be explained by appeal to external factors acting upon a system—it it not connected with meeting any needs of the system. A feather in the wind moves but does not behave, since there is no way in which the feather contributes to being airborne in an ongoing fashion, nor does being airborne contribute to the needs of the feather.

With this in mind, we can propose understanding cognitive behaviors as those that support sensorimotor coupling and that serve a function for the system, at least for the purposes of evaluating minimal models. We do not propose this as an all-purpose definition of cognition, but as a workable one that draws from a contemporary empirical research program and serves the evaluative purposes of this paper. It is a biologically-grounded functional definition in the sense that cognitive processes and behaviors are those that involve aiming to achieve its existential goals (Lyon et al., 2021) and can

---

and proposals for determining explanans). We thank the reviewer for urging us to clear up this misunderstanding.

[4] We refer again here to the various unresolved controversies over (for example) the necessity of representations (Fodor, 1975) and/or language (Davidson, 1975) for fully-fledged cognition, the possibility for an extended constitution of minds (Adams, 2018; Adams & Aizawa, 2008), and the direct or mediated operation of perception-action (Fodor & Pylyshyn, 1981).

be investigated empirically in various ways. To summarize, rather than advocating for any specific framework for cognition, we use a functional definition of cognition as sensorimotor coupling aimed at meeting the existential needs of the system, and with no commitments about what underpinnings facilitate these processes. Taking this approach to the essential features of cognition, does a similarity between the behaviors of a cognitive system and the behaviors of an active material establish that the latter can be a minimal model of the former? In order to argue that this is the case, we would first have to establish what kinds of behaviors qualify as those exhibiting the essential features of cognition as described above. In the following two sections, we evaluate what kinds of organization or activities might lead researchers to think that the activities of a kind of active material, motile oil droplets, could make them minimal models of cognition.

3 Minimal models of cognitive organization?

One type of active material that may be thought to engage in cognitively interesting behaviors is a self-catalyzing oil droplet (Hanczyc, 2011). Motile oil droplets form when oil is placed into a mixture of water and surfactant. If a Marangoni instability occurs (a flow along the interface due to an imbalance in surface tension), the droplet can form a convective cell. In repeated studies on nitrobenzene oil in a high alkaline solution (pH = 12), Martin Hanczyc and others (Hanczyc & Ikegami, 2010; Horibe et al., 2011; Hanczyc, 2014; Čejková et al., 2017) have been able to induce a kind of chemotactic process with these motile droplets using oleic anhydrine, which converts to oleic acid on contact with water. Due to this reaction and accompanying waste dispersion, the droplets will travel through their liquid environment autonomously (Marchant, 2011).

This is possible because of the self-organization of the chemicals into the convective cell, a process which Hanczyc and Ikegami (2010) compare to the autopoietic organization of living cells (Maturana & Varela, 1980). For a system to be autopoietic, as all living systems are, they must be self-producing and self-maintaining, autonomously seeking out energy from their environment in order to maintain their own boundary and identity. The mutual processes of selective openness to the environment and boundary maintenance are seen by Maturana & Varela (and subsequently Hanczyc & Ikegami) as providing a good reason to believe that all living beings must be cognitive, since they are required to be selective about when to be closed to the environment (to avoid dissipation or harmful elements) and open to the environment (to take in energy), which we can observe through their sensorimotor behaviors.

Because of the convection within the cell, new fuel supplies are continuously brought to the surface of the droplet, meaning it can maintain motility until the fuel is exhausted. However, if fuel is put into the surrounding medium, the droplets can also absorb new supplies of fuel. For living systems, being able to obtain energy from the environment is a necessity for avoiding equilibrium and is a result of millions of years of co-evolution between organisms and their environment, so as not completely determined by the organism itself. Hanczyc and Ikegami point out that a similar co-determining system-environment relation is also the case for their motile droplets:

"Even after the autonomous droplet emerges, it is still controlled by the environment and its own temporal changes. This is what we consider to be the congruent regularity of the droplet motion, which is the product of both droplet and environment" (Hanczyc & Ikegami, 2010, p. 238).

We might think, then, that for those committed to the autopoietic framework, the self-organization and self-maintaining activity of the droplets might make them minimal models of cognition. However, for autopoietic theorists, it is not simply the organization, but the *way* it generates organism-environment relations that is key. Though autopoietic theorists tend to eschew functionalist language for various reasons (see Di Paolo et al., 2017), it is nonetheless important that the behaviors of the system be in service of self-maintenance. For example, one of the key features of an autopoietic system is its selectivity. This is why, for instance, Benard convection cells are not autopoietic, even though they are self-organizing and are often said to have an autopoietic structure (Iniguez, 2001; Collier, 2004). Moreover, autopoietic selectivity is guided by the existential needs of the autopoietic system: damaging features of the environment are (to the best of the system's ability) avoided, waste products are allowed passage out, and things that can be consumed for energy are allowed in. We find no such selectivity in service of self-maintenance in the behavior of motile oil droplets.

As Batterman and Rice (2014) argue, minimal models are explanatory in virtue of their ability to instantiate the essential features of a macrobehavior belonging to a universality class. A structure, though, is not a behavior. We therefore need more than structural resemblance to establish the essential features of cognition which a minimal model could instantiate. Fortunately, there are a multitude of examples of active droplets engaging in behaviors similar to those behaviors we see in cognitive systems. One of the most frequently referenced behaviors is that of gradient-following, which resembles chemotaxis in bacteria. In motile oil droplets, the oleic anhydrine to acid reaction lowers the pH at the droplet/medium interface, so a droplet can follow a pH gradient, and "is therefore capable of chemotaxis as found previously only in living systems" (Hanczyc, 2014, p. 1041). Because the oil droplet's motion is generated by convection internal to the system, and because of the ability of the droplet to follow a pH gradient, Hanczyc argues "that the droplet has an interface that can sense its local chemical environment and an internal convective flow acting as a motor. Therefore, the system possesses a primitive form of sensory–motor coupling" (Hanczyc, 2011, p. 2886; Hanczyc & Ikegami, 2010). The motor is considered to be the convection which brings new fuel to the interface, and the droplet senses local changes in pH by means of an interface imbalance. The gradient-following behavior even allows any droplets capable of following a pH gradient to find the shortest path through a maze, given that a chemoattractant has been placed at the end. The chemoattractant diffuses into the medium, and the droplet will follow the strongest gradient (shortest path) through the maze (Lagzi et al., 2010).

Now we seem to have a clear case of a behavior oftentimes associated with cognition, chemotaxis, but what determines that we ought to think of the droplet's gradient-following as falling under the universality class of *cognitive* behavior? In bacteria, the function of chemotaxis is to link the metabolic needs of a given bacterium with the external environment. Droplet chemotaxis is not likewise fulfilling any existential needs of the oil droplet system. The chemoattractant is not attractive *because* it is an energy source, but because of a chemical reaction that in no way contributes to the maintenance of the system. This point is highlighted by Bich and Moreno (2016) in their discussion of why regulation is a necessary component of a cognitive system. They explain that while self-propelled oil droplets exhibit gradient-following behaviors, these behaviors are unregulated from the perspective of the system itself. Though the convection dynamics cause the movement, the environment alone determines the directionality of the system, which is qualitatively different from bacterial chemotaxis in a way that distances it from anything we might consider cognitive:

> "[T]he direction of droplets taxis is directly controlled by external conditions (pH gradients). It is a very interesting case of physical dynamic stability, realised through the direct coupling between the droplet dynamics and its environmental gradients. While droplets's movement is to an important degree co-determined by the environment, and its direction is governed by the external gradient, this is not the case for bacteria, which are 'intrinsically active' (Bechtel, 2008) due to their capability of self-maintenance. Their movement, in fact, is internally generated by the same regime that produces and maintains the system, and it is also 'inherently goal-oriented' (Barham, 2012). Using Kauffman's expression, bacteria are autonomous systems because they 'act on their own behalf' (Kauffman, 2000)" (Bich & Moreno, 2016, p. 16)

So, while self-propelled oil droplets exhibit a similarity to bacterial chemotaxis, which also involves gradient-following, this gradient-following does not involve meeting the existential needs of the system. Or, in more autopoietic terms, the "intrinsic goal that produces the movement is the maintenance of the organism" (Bich & Moreno, 2016, p. 16) in the case of bacterial chemotaxis, but not in the case of the oil droplet. The droplets might offer minimal models of gradient-following activity, but the leap from this to being a minimal model of *cognition* seems like a big one.

Egbert and Di Paolo (2009) have made the broader point that a lack of behavior is also a problem for computational models of autopoiesis, which they say can self-maintain but not behave:

> "Generally the simulated autopoietic entities exist in an environment which requires no organism-scale action to continue to exist (e.g. McMullin 2004; Varela et al., 1974). A few more recent models have demonstrated agents performing a slightly extended autopoiesis; extensions such as incorporating a simple behavior such as osmotic crisis avoidance (Ruiz-Mirazo & Mavelli, 2007) or chemotaxis (Suzuki & Ikegami, 2009). In these cases, the added behaviors are actually extensions of the mechanisms of autopoiesis—they are inseparable from the autopoiesis. To stop the mechanism of behavior is to stop the mechanism of autopoiesis. However, this is not the case for the majority of behaviors observed in nature that stop and start while autopoiesis continues." (Egbert & Di Paolo, 2009, p. 388).

In other words, the activity that we find generated in models of autopoiesis is an extension of autopoietic processes; stop the behavior, and autopoiesis itself stops. The behaviors do not involve the kind of selectivity required for the maintenance of an autopoietic system; they are merely contingently attached to these autopoietic processes. This is not the case for living systems, for which autopoietic organization and the behavior of the system can be distinguished, though they are mutually supporting.

This is noted in the more recent work of Egbert (2021) on the self-preserving behaviors of oil droplets. In evaluating the possible metabolic basis of the activity of these droplets, he is clear to specify that it is the chemical reactions “taking place on the surface of the droplet” that cause it to follow gradients “towards conditions that facilitate or extend the life of those very same reactions” (Egbert, 2021, p. 2). Where bacterial chemotaxis relies on assessment through the organism’s metabolism, in which case the metabolic needs of the systems underlie its movement, the movement of the oil droplet has no whole-system regulatory aspect. An oil droplet can not be said to be undertaking metabolism-dependent movement, so Egbert specifically points to the dissipative system comprised of “(i) the chemistry on the surface of the droplet; (ii) the marangoni flow that it creates; and (iii) motility driven by the marangoni flow” as possible engaging in a viability-based response to existential needs (ibid., p. 4). Egbert links this to minimal forms of cognition:

> The rate of hydrolysis increases with alkalinity, and Marangoni flow is such that the reaction drives the droplet toward local environmental conditions that accelerate that reaction. This is an interesting example of what has been called metabolism-based behaviour—a precarious dissipative structure which regulates its environment in response to its own metabolic health (Egbert et al., 2010 [1]; Egbert & Pérez-Mercader, 2016) which has been considered a basic form of cognition…

Elsewhere, Egbert also makes this connection, saying that motile oil droplets may be “minimal examples of *autonomous agents–c*onstrued as precarious self-maintaining systems that act to satisfy their own needs” (Egbert, 2020, p. 1). Note that because of commitments to the autopoietic framework, the definition of autonomous agency is very close to the functional account of cognition we’ve provided.

The claim that motile oil droplets might be minimal examples of autonomous agents might be true, if, in conflict with the autopoietic account, agency and cognition are construed as separate phenomena–but that would also be on the condition that acting to satisfy the system’s needs could be purely accidental. There is no evidence that oil droplets act in a way that actually does contribute to the maintenance of the system unless the droplet is following a course explicitly set for such a purpose by an outside force. As Egbert has pointed out, the activity of motile oil droplets is generated purely by reaction at the boundary of the Marangoni effect rather than being generated by the whole system, and without aiming to fulfill the existential needs of the droplet as an autopoietic system. Likewise, we have no reason to think that other kinds of active materials are not also acting in ways that are purely accidentally contributing to or harming their well-being.

In this section, we have argued against the idea that there is a general universality class of cognitive behaviors, established by organization and system-generated norms, to which active materials may belong. Still, motile oil droplets do seem able to provide possible minimal models of specific types of behaviors that, for some systems, would be considered cognitive when they are furthering the existential goals of the system, as is the case with gradient-following and chemotaxis. This would not qualify them as minimal models in the sense Batterman and Rice propose, as they do not have explanatory power in virtue of belonging to the universality class of cognition or cognitive behaviors. Another option, then, is that rather than being minimal models of cognition *proper*, they provide models of some particular cognitive capacities (e.g. memory, navigation). That is, instead of thinking of their behaviors as belonging in a universality class, oil droplets may provide minimal models explanations of specific cognitive capacities. We evaluate this possibility in the next section.

4 Minimal models of cognitive capacities?

We have raised some concerns with the idea that active materials are minimal models of cognition in the explanatory sense proposed by Batterman and Rice (2014), where both model and target system belong to the same universality class. Using motile oil droplets as our example, we have shown that they do not exhibit sensorimotor coordination in behaviors serving the integrated existential needs of the system. If cognition itself is not being modeled directly, this may still leave open the possibility that particular *cognitive capacities* are being modeled, such as “sensing, information processing, memory, valence, decision making, learning, anticipation, problem solving, generalization and goal directedness” to again cite the basal cognition toolkit (Lyon et al., 2021). If minimal models are expected to pick out only one particular capacity of a cognitive system, then it might seem that active materials could display the essential features of just that single capacity.

For example, the motile oil droplet is hypothesized to have several ways to engage in sensory-motor coupling, where the coupling mode changes depending on the context. For droplets, these modes are flexible because there are both internal factors (convection flow pattern inertia and internal distribution of chemicals) and external factors (trails and accumulation of chemicals), both of which can be thought of as memory dynamics (Ikegami et al., 2015). As the droplets expel chemical waste, they leave trails through their medium that, as they continue to follow the chemical gradients in the system, affect their movement. This has been proposed to be a primitive kind of memory because it functions as an external device that alters the ongoing behavior of the system through chemical changes the system itself has made in the environment (Hanczyc & Ikegami, 2010; Ikegami et al., 2015). These changes can affect the behavior of the system, not just its movement trajectory, by altering the modes of sensori-motor coupling which are dominant at a given time (e.g. inertial dynamics or gradient-following and avoidance of waste trails).

Let’s compare this to an organism that the oil droplet might be thought of as modeling: the slime mould *Physarum Polycephalum*. Slime moulds are acellular organisms that produce extracellular trails as they traverse their environments, which alter their future movement. They avoid crossing these extracellular trails, as they mark areas in which the slime mould has already recently foraged (Smith-Ferguson & Beekman, 2020). Endorsing Baluška and Levin's (2016) definition of memory as "experience-dependent modification of internal structure, in a stimulus specific manner that alters the way the system will respond to a stimulus in the future as a function of its past" (p. 902), Sims and Kiverstein argue that we should consider the slime mould's production of extracellular slime as also fitting this definition, though external rather than internal to the system, in a "generalised biological memory" (2022, p. 1). They justify this through reference to the extended mind hypothesis (Clark & Chalmers, 1998), which proposes that we should consider external memory devices–specifically those created by the system itself–as constitutive of cognitive systems in the case that these devices make a functional contribution to the operation of that particular cognitive capacity. The internal/external divide is irrelevant to the operation of memory itself, as these differences do not matter functionally in the way that they shape the system's behavior. This is in line with researching scientists' approach to understanding the function of extracellular slime (Smith-Ferguson & Beekman, 2020; Reid et al., 2012).

As we've argued above, resemblance relations alone cannot establish that the behaviors of these different systems belong in the same universality class. The behavioral similarities are clear: both systems leave trails which influence their own future movements through their environment. However, when the slime mould leaves its trails of extracellular slime, this behaviour signals something to the system that it is able to act on depending on its existential needs. Slime moulds feed via extending and contracting their body to slowly "crawl" over large areas, avoiding danger (for instance, bright sunlight or harmful chemicals like salt) and efficiently searching their environment for nutrients before moving on to a new locale (Nakagaki et al., 2004). Ruling out areas that have already been searched is necessary for the slime mould to conserve energy for further exploration; though this is described by researchers as a *choice* "because when no previously unexplored territory is available, the slime mold no longer avoids extracellular slime" (Reid et al., 2012, p. 17,490). For slime moulds, then, external memory is used to conserve the energy of the system by marking previous paths, and in a way that is sensitive to (and driven by) the existential needs of the system.

The memory dynamics observed in oil droplets are behaviorally similar to the extracellular memory of the slime mould, but without being flexible to the needs of the system. We know that the oil droplet's behavior is not guided by metabolic needs, so the avoidance of trails serves no purpose for the system–it is a contingent facet of ongoing chemical reactions (Ikegami et al., 2015). The oil droplets studied by Ikegami et al. (2015) do not decouple themselves from this environmental signal when context demands it, unlike the slime mould. We can see a clear difference between a mould and a droplet here: the slime mould's behaviours are part of (and to some extent subordinate to) a broader scheme of self-maintenance, while the behaviour of oil droplets serves no function for the system.

Again, we want to stress that researchers working on motile oil droplets do not say that these systems are cognitive, nor do they claim that the explanatory value of using these systems as minimal models comes from belonging in the same universality class as cognitive systems (or capacities). In fact, Ikegami et al. (2015) are clear about what they see as being the fundamental differences between cognitive (living) and non-cognitive (non-living) systems: "Intentional movement does not make sense on the molecular or chemical level but on the cognitive level where we can investigate the properties of [sensori-motor contingency] selection." (p. 353). Sensori-motor contingency selection cannot be thought of as cognitive, nor as minimally modeling cognition, without the system having existential needs which the selection serves to fulfill–otherwise there simply is no *selection* happening that makes use of dynamics or processes internal to the system, just forces external to the system acting upon it with differing gravity. We may have a minimal model of a behavior, but it is not a cognitive behavior in both instances and therefore does not establish a universality class of the latter type.

What we have argued so far demonstrates that these active materials cannot be thought of as minimal models of cognition, nor of cognitive behaviors, in the way proposed by Batterman and Rice (2014). If these systems are in fact valuable as minimal models of cognition or of cognitive processes in some other way, then there needs to be some other way to understand their explanatory power. This will be proposed in the next section, but before that, there is at least one further possible way of understanding the claim that active materials can be minimal models of cognition, one that uses the explanatory strategy employed as its justification. After discussion of the resemblance between the behavior of motile oil droplets and the behaviors of cognitive systems, Hanczyc and Ikegami ask a set of questions about the necessity of explanations that involve positing cognition: "So the question is, how can we derive sophisticated intelligence from a merely thermodynamic system? In other words, when is it necessary to use the intentional stance (Dennett, 1987) to describe a system's behavior—for example, by using sensing, or cognition, instead of reaction, or hysteresis?" (Hanczyc & Ikegami, 2010, p. 233). The intentional stance derived from Dennett (1987) involves using language usually reserved for cognizing systems (perceiving, acting, thinking, believing, knowing, etc.) when necessary for crafting explanations about said systems (though it need not involve any real attribution of cognition to these systems). We gain no explanatory value from describing a rock as *desiring* to roll down a hill; we use the language of the physical sciences to explain its descent. Explaining the activities of a bird, however, might require the language of *desiring* food, *perceiving* danger, and so on.

While adopting the intentional stance when thinking about active materials is useful, it is important not to read too much into this investigative maneuver. Adopting the intentional stance or using intentional language is not a response to findings, where we discover that a system is cognitive and thereafter decide to use the appropriate language when describing it. Rather, it is an acknowledgement (and codification) of a series of assumptions and, to borrow Dennett's wording, *hunches* about whether we should try using cognition-laden language to gain purchase on understanding a system. In the case of oil droplet memory trails, the use of intentional concepts like memory is clearly useful to researchers, but should not be taken as a claim that a cognitive *explanation* is needed for the behavior. Using cognition-associated words as a heuristic and using them in a way that requires a cognitive explanation are two different things that should not be conflated.

Further on this point, Batterman and Rice's (2014) account specifies that its explanatory target is the macrobehavior of a system, via appeal to the ineliminable underlying features that produce that behavior. When observing fluid flow it is fairly straightforward that the macrobehavior of flow is underpinned by a fluid possessing certain features. Physics has a good handle on what counts and does not count as a fluid. So, in Batterman and Rice's account, the claim that the LGA model explains fluid flow transitions into the understanding that this tells us something about fluids themselves–the LGA model allows us "to investigate and to understand the actual behavior of *real fluids"* (Batterman & Rice, 2014, pg. 359, emphasis added).

It is not so easy to do the same for *real cognition*. The elision from macrobehavior to cognitive behavior is a trickier business. The sorts of behaviors we might be inclined to approach via the intentional stance seem quite different in this respect to Batterman and Rice's (2014) examples. While an explanation of fluid flow comfortably transitions into a better understanding of fluids (whatever the actual physical implementation of this may be), cognitive behaviors are a motley assortment of macrobehaviors without essential features we can explain through appeal to underlying features. Cognitive behaviors are not obviously "made of" anything in the way that systems exhibiting fluid flow are. The pursuit of existential needs does not have essential components playing the same functional role wherever we find a type of behavior.

Therefore, the notion that minimal models might reveal to us what cognition is by demonstrating what its minimal features are is a non-starter. In order to move from intentional macrobehavior to cognitive capacity (or cognition), there needs to be a more substantial theoretical framework in place that identifies some or another substrate or vehicle as the only possible explanans for cognition *and* interprets relationships between specific cognitive behaviors and cognitive capacities. The problem is that taking this step circumvents establishing the function of that behavior *for* the system, which we have argued above is needed to establish that the behavior is cognitive in the first place. This is because such a move preemptively limits the study of cognition to focusing on systems that satisfy requirements on what the underpinnings of cognition are assumed to be on a theoretical basis rather than focusing more practically on behaviors which serve a functional role for organisms.

Batterman and Rice's (2014) account of the explanatory justification of minimal models cannot apply to minimal models of cognition, because there simply is no universality class of either cognition or cognitive behaviors that meets the same "minimal model" explanatory criteria as fluid flow. This also dissolves McGivern's dilemma (2019): if a system is a useful minimal model of cognition, then it is also cognitive itself via the universality class; but if that system is not itself cognitive, then it is unclear why it is useful as a minimal model of cognition. However, in rejecting the universality class view of minimal model explanations of cognition, we do not reject the idea that minimal models such as active materials can be explanatorily useful. In the section below, we will go into more detail on why it is that minimal models of cognitively interesting behaviors can still tell us a good deal about cognition, and are explanatorily justified.

5 Minimally cognitively interesting agents

According to Batterman and Rice (2014), minimal models have explanatory power in virtue of belonging in the same universality class as the target phenomena. The previous sections have shown that the behaviors of active materials do not belong in the same universality class as the cognitive behaviors of biological systems by virtue of their organizational structure or by resemblance relations to specific cognitive capacities. Further, we have suggested that the idea of a universality class that applies to all or even a large class of cognitive behaviors is unlikely given the role that behaviors must be playing *for* individual systems to be considered cognitive.[5] Per the minimal models account of explanation, if the behaviors of active materials are not cognitive, then it is not clear how they can offer explanatory value towards understanding cognition. However, there is another way of thinking of the value of these models, one which offers explanatory value through an alternative methodological approach. We now go to Beer's *methodology of minimally cognitive behaviors* to demonstrate that minimal models can still help us understand and explain cognition, though these explanations are not grounded in membership in a common universality class. Instead, Beer's models provide *dynamical explanations*. Understanding Beer's methodology and how it is used to provide explanations will shed light on how active materials models can also have explanatory value despite meeting neither universality class nor standard mechanistic modeling criteria.

The term 'minimal cognition' has often been adopted as shorthand for work on minimally cognitive agents, as popularized by Beer (1996, 2020a, b) and colleagues. Beer's framework originated as a response to Clark and Toribio's (1994) claim that the domain of interest to cognitive science was that which tackled "sufficiently 'representation-hungry'" problems (p. 418), or those that traditionally seemed to require positing a representational entity in their explanation (language, group coordination, memory, etc). The program developed by Beer, drawing on contemporaries like van Gelder (1995) and Brooks (1991), instead shows that small neural networks can often evoke "the simplest behavior that raises cognitively interesting issues" (Beer, 1996, p. 422) without need of representations. Models used by Beer and others often involve resources from an expanded cognitive science toolkit (dynamical systems theory, embodiment, cognitive offloading, robotics, information theory, and so on) to achieve behaviors without complex computational or representational processing within the system.

However, following Pamela Lyon (2020), he warns about the drifting usage of 'minimal cognition' from a methodological program to an ontological category. That is, the terminology was not intended to imply that the models instantiate the essential

---

[5] We are grateful to an anonymous reviewer for pointing out that our use of "universality class" in the technical sense here does not exclude the possibility there may be other ways of grouping cognitive agents or some particular types of behaviors undertaken by some of these agents. We reference above the universality class example of fluid flow used by Batterman and Rice (2014), which is a behavior of fluid achieved at multiple scales and involves the essential features of locality, conservation, and symmetry. Our argument here is strictly in relation to this point; there is no such class of cognitive *behaviors* produced by some identifiable, isolatable essential features *of the behaviour* in each case and across all scales in the same sense.

features of cognition, that the models point to the essential features of cognition in target systems, or that 'minimal cognition' implies a criteria for inclusion. Treating minimal cognition as an ontological category gets away from the initial spirit of Beer's project, as it was not meant to define cognition in any particular way, establish boundaries, or "to propose specific criteria that demarcate the cognitive from the noncognitive" (Beer, 2020a, b, p. 3). The methodology of minimally cognitive behaviors, hereafter MMCB to avoid this confusion, is intended to provide a concrete methodological means of investigating and predicting behavior, not establishing the ontological prerequisites for cognition as specified by high-level theories and overarching frameworks.

So what makes a system's behaviour cognitively interesting in the eyes of the MMCB? In Beer's early work in this area (1996), he demonstrated that a simple evolved two-dimensional model agent with an eye (array of distance sensors), two motors, an "arm", and a "hand" can engage in a number of activities. The agent, controlled by a dynamical neural network and simple feedforward circuits, but without any representational capacities, could orient itself around, navigate, and distinguish between objects (circle/square) in a simple sorting task. It could sense which gaps between objects it could fit through and which it could not. The agent could also manipulate objects and build simple structures. It was supposed that the agent might be able to engage in simple cooperative tasks with other agents of its kind as well. Knowing that these systems are not employing internal symbol-manipulation to achieve these tasks warrants further investigation–they are doing something cognitively interesting and the idea is that anyone interested in cognition would likely want to better understand *how* these tasks are possible. For example, circle/diamond discrimination in the simple 2D models leads Beer to ask "Can we identify 'circle' and 'diamond' (or 'smooth' and 'pointy') detectors in these circuits?" (1996, p. 8), where these did not involve the representational capacities argued by Clark and Toribio (1994) to be the lone resource and focal point of cognitive science (Beer, 2020a, b).

This ongoing work has led to the development of a research program for analyzing the capacities of minimally cognitively interesting model systems and positing empirically testable alternatives to representational explanations within the purview of the cognitive sciences. These models are explanatorily important in a general sense because they can demonstrate what may *not* be needed in an explanation of a particular behavior. That is, they show how assumptions about common underpinnings for cognitive behaviors might be wrong, or their necessity overinflated. This allows us to narrow down and refine our understanding of the internal dynamics or mechanisms necessary for a target system to engage in a behavior, or to demonstrate how dynamics external to the system can contribute to a behavior, while facilitating interpretation and comparison between competing frameworks invoked in the explanations of those behaviors. The models help "to minimize the impact of our a priori preconceptions about how an agent that successfully performs a given task ought to work, thus making the analysis of how it actually does work an interesting and insightful exercise" (Beer & Williams 2015, p. 3). In this way, these models can be explanatorily valuable by revealing ways in which cognitive science has built in assumptions about what *will* be needed in an explanation of behavior.

More important to the argument here is that these models also themselves provide explanations of behaviors. The MMCB has been a driving force in advancing the idea of *dynamical explanations.* In dynamical explanations, "the explanatory focus is on the structure of the space of possible trajectories and the internal and external forces that shape the particular trajectory that unfolds over time, rather than on the physical nature of the underlying mechanisms that instantiate this dynamics" (Beer, 2000, p. 96). A dynamical explanation utilizes dynamical systems theory, a mathematical theory that interprets behaviors through differential equations that describe the agent-environment system's state over time.

Again, the MMCB involves a model system that is shaped through evolutionary algorithms in a specific environment, so these explanations take up behaviors as a trajectory involving dynamics of agent-environment coupling (Beer, 2003). In this way, Beer's models do not describe the workings of any specific cognitive mechanisms found in a target system. Specific implementations in real organisms are not the explanatory target of the MMCB. This aligns with the fundamental motivation of the basal cognition approach; it is concerned with finding out how organisms accomplish the same vital functions common across the tree of life, though implemented in dramatically different ways. The object of explanation for models used in MMCB are not often target organisms; for these models, the targets are more frequently putative cognitive phenomena (such as the sorting of circle/square mentioned above), or cognitive concepts or definitions.

For example, Beer has utilized gliders in the Game of Life (2004) to examine and compare framework-specific explanations of cognition that have their basis in the autonomous self-maintenance of a system. Beer treats gliders (2004, 2014, 2018, 2020a) as autopoietic entities, based on the biological theory of Maturana and Varela (1980) which asserts that self-production and self-individuation are the hallmarks of life and cognition. In this framework, the ability of an entity to withstand perturbations without disintegrating is possible because of the entity's autopoietic processes (self-production over time), and this domain of interactions is considered its "cognitive domain." Beer demonstrates that gliders offer a minimal model of this process, as they can withstand perturbations based on previous states of the system much in the same way other autopoietic entities can. For instance, Beer offers examples of two identical gliders that undergo a series of perturbations (Beer, 2004). Though both start with the same structure, the series of perturbations are different for each, leaving them in different states when they encounter the next perturbation. On encountering the *same* perturbation after encountering *different* perturbations, each glider is differently prepared through modifications resulting from the first encounter, leading one to be able to withstand the second perturbation while the other is destroyed. Thus, Beer shows that "each perturbation that a unity experiences, as well as the structural changes that it undergoes even in the absence of perturbations, influences its sensitivity and response to subsequent perturbations" (Beer, 2004, p. 316).

Beer offers a series of such examples in order to demonstrate that gliders can instantiate and test many of the ideas of autopoietic theory in a concrete model which can be used to map out a single entity's cognitive domain (Beer, 2014) as determined through this specific theoretical framework. These models can be used to explain how autopoietic processes function for the self-maintenance of a system in a particular environment, though implementation of these processes would look much different for other types of systems, and offers neither a universality class nor causal explanation. Beer does not make any claims about the implications of this for what we might attribute to the model itself, but he is clear that it does not demonstrate anything along the lines of what it would be to *satisfy the essential features of cognition* (Beer, 2020b; Lyon, 2020). And as with the active materials above, these minimal models are in no way said to be cognitive themselves: "Note that the intent here is not to argue that such patterns are "really" alive or that they capture all relevant characteristics of living systems. Rather, I wish merely to utilize GoL as a simple model universe in which we can explore in detail the many issues raised when attempting to apply the definition of autopoiesis to a concrete system" (Beer, 2020a, b, p. 3).

The MMCB can be used to test the claims and concepts of a particular framework, but it does not tell us about the essential features of cognition–nor does that model's behaviors or organization tell us that the model itself exhibits those essential features. The MMCB shows us what interesting behaviors can be achieved without the use of representational components, for example, but further steps would be needed to say that these behaviors are indicative of the essential features of cognition. Beer has said that it would be a mistake to think that his intention was to model *cognition*, rather than offering a testable model of a putative cognitive phenomenon (e.g. perturbation dynamics, structural coupling) (Beer, 2020a, b). However, it does not provide a basis for claiming that a behavior being replicated within a model is a cognitive behavior in all systems. As we've argued above, that would require demonstrating that in each case, for each system, that behavior is functionally aimed towards meeting the existential needs of that very system (even though the sensorimotor activity might be instantiated through quite different components altogether).

There may be some concern that these types of dynamical explanations do not genuinely *explain*. As we have already shown, they do not establish membership in a universality class in the vein of Batterman and Rice's minimal model explanations. Likewise, they do not fall into the class of another well-established variety of scientific explanation, where the objective is to establish causal relationships through mechanistic models that explain the complex causal organization that produces a given biological or cognitive phenomenon, and in so doing provide handles for control (Craver, 2007). Instead of this kind of causal or mechanistic reasoning, MMCB aims to provide a mathematical lens which can be applied in rigorously testing cognitive phenomena and concepts.

In this way, the MMCB only purports to provide one of many possible explanations of a behavior, though reasons for preferring dynamical explanations in some situations are provided (see e.g. Beer, 1995; Beer & Williams, 2015). Likewise, Batterman and Rice do not argue for the conclusion that their minimal model explanations are the only way that minimal models can explain. While some mechanists have proposed that explanations should be limited to those that are causal, and even then, only causal in a way that pinpoints the underlying mechanism across both model and target system (Craver & Kaplan, 2011, Kaplan & Craver, 2011), there are several recent papers demonstrating that dynamical explanations can indeed be causal, and debating this further would be out of the scope of our argument (see Van Eck, 2018; Meyer, 2020a, b). We leave it an open question whether the cognitive phenomena and concepts explored in the MMCB require causal-mechanical explanations despite the existence of this highly successful research program that utilizes alternatives.

The MMCB demonstrates how minimals models of cognitively interesting behaviors can have explanatory power, even though these minimal models do not belong in the same universality class as genuinely cognitive systems. Establishing the reasoning at work in Beer's methodology allows us to characterize a more general set of explanatory norms that can also be applied to the active materials we have shown don't fall under Batterman and Rice's (2014) explanatory account. Active materials provide a testing ground for cognitive phenomena and concepts through a chemical lens, and modeling of them likewise provides the potential for a mathematical lens that can be used to explain how various putative phenomena fit into cognitive frameworks and to empirically test those phenomena. Importantly, in all of the models discussed, agent-environment coupling is key for understanding behaviors. Where Beer's models use an evolutionary algorithm through which model systems are able to undertake interesting behaviors after many iterations, the activities of motile oil droplets require precise engineering of both the droplet and its medium to evoke behaviors.

Conversely, looking strictly at mechanisms internal to the system, or *essential features* of either the droplet or its behavior, do not provide much in terms of an explanation of behaviors. Understanding the motion of a droplet engaging in chemotaxis, for example, involve a mathematical analysis of the way environmental conditions support ongoing chemical reactions that enable the droplet to persist (Egbert, 2020). Such models do not purport to offer general explanations of chemotaxis, and they do not purport to offer explanations applicable to a specific biological target system. However, they can explain by demonstrating how particular dynamics of agent-environment coupling might support basic metabolism-based behavior, for instance (Egbert, 2020). When we move from the domain of models to that of real organisms, we can investigate how both bacteria and eukaryotes, via very different mechanisms, might be supported by similar agent-environment dynamics in their own processes of chemotaxis.

The reasoning behind the MMCB might seem to be in line with some of what Batterman and Rice have said about establishing the essential features of a behavior: "[Minimal] models are explanatory in virtue of [there] being a story about why large classes of features are irrelevant to the explanandum phenomena" (2014, p. 356). We don't disagree, as this would certainly be the case for a large number of phenomena that are the domain of other sciences, such as physics and chemistry. But our argument holds against the possibility of there being an explanandum phenomena in the case of cognition that instantiates as a behavior across the universality class of cognitive systems. Rather than taking the idea that active materials can be 'minimal models of cognition' off the table, Beer's methodology shows that the explanatory power of minimal models is not limited to causal-mechanistic models or the types of explanations employed in physics. As with the MMCB, to understand and explore what active materials can help us learn about cognition might require us to get comfortable using a richer variety of explanatory resources. Cognition being the complex phenomena that it is, embracing a wide number of explanatory possibilities rather than preemptively excluding some of these resources based on a narrow conception of explanation seems like the best way to make interdisciplinary progress.

6 Conclusion

We've shown that the account of the explanatory power of minimal models, as provided by Batterman and Rice (2014), does not work for cognition. On the functional understanding that cognitive behaviors involve sensorimotor activity in support of the existential needs of a system (drawn from Lyon et al., 2021), cognition is not indicated simply in a behavior's tokening of a type. This rules out the idea that cognitive behaviors have a universality class. It should be no surprise that minimal model explanations, like virtually any given explanatory strategy, have their limits. Taking a survey of historical and contemporary accounts of explanation - covering law, mechanistic, unificatory, for instance - it is often the case that there are areas where they thrive, and others where they get no traction. Though a popular option, mechanistic models may explain well in some sub-domains of biology and cognitive science, but poorly in others (Meyer, 2020b). Minimal model explanations are no exception here. While models dealing with physical and chemical, or even neuroscientific phenomena (Ross, 2015) may be well-treated by a minimal model explanation per Batterman and Rice (2014), the idiosyncrasies of the subject matter of cognitive science constrains its use there.

However, the use of active materials as minimal models to study cognition can be explanatorily justified in ways other than through appeal to a universality class. The MMCB of Beer (1996, 2020a, b), shows how minimal models can help explain cognition through the engineering of simple sensorimotor dynamics that facilitate theory testing and comparison of framework-specific explanations. Active materials, our target in this paper, similarly engage in cognitively interesting behaviors even though they are not themselves cognitive in any sense we are interested in here. What active materials represent for scientists and philosophers is an almost unique situation: systems that exhibit cognitively interesting behaviours but in the absence of cognition. Here new interventions in the spirit of Beer's work on autonomous artificial agents and Game of Life automata can (and are) being developed. Where this will take us in our understanding of cognition, and which frameworks will enter the picture to scaffold this work, remains to be seen. What is beyond doubt is that active materials provide a promising avenue for enhancing that understanding.

Acknowledgements The authors would like to thank Patrick McGivern, Matt Sims, Caroline Stankozi, Fred Keijzer, and Alexander Hölken for comments and discussion, as well as two anonymous reviewers for their helpful feedback.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. NB received support through the Templeton World Charity Foundation Grant: "Intelligent Agency on Multiple Scales" (TWCF0463) and a University of Wollongong PERL Fellowship award.

Declarations

Conflicts of interest The authors declare no conflicts of interest.

References

Abramson, C. I., & Levin, M. (2021). Behaviorist approaches to investigating memory and learning: A primer for synthetic biology and bioengineering. *Communicative & Integrative Biology*, *14*, 230–247.

Adams, F. (2018). Cognition wars. *Studies in History and Philosophy of Science*, *68*, 20–30.

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

Akagi, M. (2018). Rethinking the problem of cognition. *Synthese*, *195*(8), 3547–3570.

Allen, C. (2017). On (not) defining cognition. *Synthese*, *194*(11), 4233–4249.

Baluška, F., & Levin, M. (2016). On having no head: Cognition throughout Biological Systems. *Frontiers In Psychology* 7.

Barham, J. (2012). Normativity, agency, and life. *Studies In History And Philosophy Of Biological And Biomedical Sciences*, *43*, 92–103.

Batterman, R. W., & Rice, C. C. (2014). Minimal model explanations. *Philosophy Of Science*, *81*, 349–376.

Bechtel, W. (2008). *Mental Mechanisms: Philosophical perspectives on cognitive neuroscience*. Psychology Press.

Bechtel, W., & Abrahamsen, A. (2006). *Phenomena and mechanisms: Putting the symbolic, connectionist, and dynamical systems debate in broader perspective*. Basil Blackwell. Contemporary debates in cognitive science.

Bechtel, W., & Abrahamsen, A. (2011). Complex biological mechanisms: Cyclic, oscillatory, and autonomous. Philosophy of complex systems. *Handbook of the Philosophy of Science*, *10*, 257–285.

Beer, R. D. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence, 72*(1–2), 173–215. https://doi.org/10.1016/0004-3702(94)00005-L

Beer, R. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In P. Maes, M. Mataric, J. Meyer, J. Pollack, S. Wilson (Eds.), *From animals to animats 4: Proceedings of the fourth international conference on simulation of adaptive behavior* (pp. 421?429). MIT Press.

Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences, 4*(3), 91–99.

Beer, R. D. (2003). The Dynamics of active categorical perception in an Evolved Model Agent. *Adaptive Behavior*, *11*(4), 209–243.

Beer, R. D. (2004). Autopoiesis and Cognition in the game of life. *Artificial Life*, *10*, 309–326.

Beer, R. D. (2014). The cognitive domain of a glider in the game of life. *Artificial Life*, *20*, 183–206.

Beer, R. D. (2018). On the origin of gliders. In *The 2018 conference on artificial life. Presented at the The 2018 conference on artificial life* (pp. 67–74). MIT Press.

Beer, R. D. (2020a). Bittorio revisited: Structural coupling in the game of life. *Adaptive Behavior*, *28*, 197–212.

Beer, R. D. (2020b). Some historical context for minimal cognition. *Adaptive Behavior*, *29*, 89?92.

Beer, R. D., & Williams, P. L., 01/2015. Information Processing and Dynamics in minimally cognitive agents. *Cognitive Science*, 39, 1–38.

Bich, L., & Moreno, A. (2016). The role of regulation in the origin and synthetic modelling of minimal cognition. *Biosystems*, *148*, 12–21.

Blair, D. F. (1995). How bacteria sense and swim. *Annual Review Of Microbiology*, *49*, 489–522.

Brancazio, N., Segundo-Ortin, M., & McGivern, P. (2020). Approaching minimal cognition: Introduction to the special issue. *Adaptive Behavior*, *28*(6), 401–405.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

Bursten, J. R. S. (2020). *Classifying and characterizing active materials*. Synthese.

Buzsáki, G. (2019). *The brain from Inside Out*. Oxford University Press.

Čejková, J., Banno, T., Hanczyc, M. M., & Štěpánek, F. (2017). *Droplets as liquid robots. Artificial Life*, *23*, 528–549.

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

Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese*, *101*, 401–431.

Colaço, D. (2022). Why studying plant cognition is valuable, even if plants aren't cognitive. *Synthese*, *200*, 453.

Collier, J. (2004). Self-organization, individuation and identity. *Revue Internationale de Philosophie*, *228*, 151–172.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the Mosaic Unity of Neuroscience*. Clarendon Press.

Craver, C. F., & Kaplan, D. M. (2011). Towards a mechanistic philosophy of neuroscience. In S. French, & J. Saatsi (Eds.), *The Continuum companion to the philosophy of science* (pp. 268?292). Continuum.

Davidson, D. (1975). Thought and talk. In S. Guttenplan (Ed.), *Mind and language*. Oxford University Press.

Dennett, D. C. (1987). *The intentional stance*. MIT Press.

Egbert, M. (2020). Marangoni based motile oil-droplets in simulated artificial chemistry. In *The 2020 conference on artificial life. Presented at the The 2020 conference on artificial life* (pp. 260–262). MIT Press.

Egbert, M. (2021). Self-preserving mechanisms in motile oil droplets: A computational model of abiological self-preservation. *Royal Society Open Science*, *8*, 210534.

Egbert, M., Barandiaran, X., & Di Paolo, E. (2010). A minimal model of metabolism-based chemotaxis. *PLoS Computational Biology, 6*(12). https://doi.org/10.1371/journal.pcbi.1001004

Egbert, M. D., & Di Paolo, E. (2009). Integrating autopoiesis and behavior: An exploration in computational chemo-ethology. *Adaptive Behavior*, *17*, 387–401.

Egbert, M. D., & Pérez-Mercader, J. (2016). Adapting to Adaptations: Behavioural strategies that are robust to mutations and other Organisational-Transformations. *Scientific Reports*, *6*, 18963.

Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1981). How direct is visual perception?: Some reflections on Gibson's "ecological approach". *Cognition*, *9*(2), 139–196.

Fulda, F. C. (2017). Natural Agency: The case of bacterial cognition. *Journal of the American Philosophical Association*, *3*, 69–90.

Gierre, R. N. (1988). *Explaining science*. Univ. Chicago Press.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy*, *21*, 725–740.

Goldenfeld, N. (2018). *Lectures on phase transitions and the Renormalization Group* (1st ed.). CRC Press.

Hagan, M. F., & Baskaran, A. (2016). Emergent self-organization in active materials. *Current Opinion in Cell Biology*, *38*, 02.

Hanczyc, M. M. (2011). Metabolism and motility in prebiotic structures. *Philosophical Transactions Of The Royal Society Of London. Series B, Biological Sciences*, *366*, 2885–2893.

Hanczyc, M. M. (2014). Droplets: Unconventional protocell model with life-like dynamics and room to grow. *Life*, *4*, 1038–1049.

Hanczyc, M. M., & Ikegami, T. (2010). Chemical basis for minimal cognition. *Artificial Life*, *16*, 233–243.

Horibe, N., Hanczyc, M. M., & Ikegami, T. (2011). Mode switching and collective behavior in Chemical Oil Droplets. *Entropy*, *13*, 709–719.

Ikegami, T., Horibe, N., & Hanczyc, M. M. (2015). Potential memory Effects in Self-Moving oil droplets. *International Journal of Unconventional Computing*, *11*, 345–355.

Iniguez, J. (2001). Rayleigh-Benard convection: A negentropic approach. Science Direct Working Paper.

Jung, K., Fabiani, F., Hoyer, E., & Lassak, J. (2018). Bacterial transmembrane signalling systems and their engineering for biosensing. *Open Biol*, *8*. https://doi.org/10.1098/rsob.180023

Kaplan, D. M., & Craver, C. F. (2011). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A mechanistic Perspective*. *Philosophy of Science*, *78*(4), 601–627. https://doi.org/10.1086/661755.

Kauffman, S. A. (2000). *Investigations*. Oxford University Press.

Lagzi, I., Soh, S., Wesson, P. J., Browne, K. P., & Grzybowski, B. A. (2010). Maze solving by chemotactic droplets. *Journal Of The American Chemical Society*, *132*, 1198–1199.

Levin, M., Keijzer, F., Lyon, P., & Arendt, D. (2021). Uncovering cognitive similarities and differences, conservation and innovation. *Philosophical Transactions of The Royal Society of London. Series B, Biological Sciences*, *376*, 20200458.

Lyon, P. (2015). The cognitive cell: Bacterial behavior reconsidered. *Frontiers In Microbiology*, *6*. https://doi.org/10.3389/fmicb.2015.00264

Lyon, P. (2020). Of what is "minimal cognition" the half-baked version? *Adaptive Behavior, 28*(6), 407–424. https://doi.org/10.1177/1059712319871360

Lyon, P., Keijzer, F., Arendt, D., & Levin, M. (2021). Reframing cognition: Getting down to biological basics. *Philosophical Transactions of The Royal Society of London. Series B, Biological Sciences*, *376*, 20190750.

Marchant, J. (2011). Oil droplets mimic early life. *Nature*. https://doi.org/10.1038/news.2011.118

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition—The realization of the living*. *Boston Studies on the Philosophy of Science*. Springer.

McGivern, P. (2019). Active materials: Minimal models of cognition? *Adaptive Behavior*, *28*, 441–451.

McMullin, B. (2004). Thirty years of computational autopoiesis: A review. *Artificial Life*, *10*, 277–295.

Meyer, R. (2020a). Dynamical causes. *Biology & Philosophy*, *35*(5), 48. https://doi.org/10.1007/s10539-020-09755-1.

Meyer, R. (2020b). The non-mechanistic option: Defending dynamical explanations. *The British Journal for the Philosophy of Science*, *71*(3), 959–985. https://doi.org/10.1093/bjps/axy034.

Nakagaki, T., Kobayashi, R., Nishiura, Y., & Ueda, T. (2004). Obtaining multiple separate food sources: Behavioural intelligence in the Physarum plasmodium. *Proceedings of the Royal Society B: Biological Sciences*, *271*, 2305–2310.

Needleman, D., & Dogic, Z. (2017). Active matter at the interface between materials science and cell biology. *Nature Reviews Materials*, *2*, 17048.

Paolo, E. D., Buhrmann, T., & Barandiaran, X. (2017). *Sensorimotor Life: An enactive proposal*. Oxford University Press.

Reid, C. R., Latty, T., Dussutour, A., & Beekman, M. (2012). Slime mold uses an externalized spatial "memory" to navigate in complex environments. *Proceedings of the National Academy of Sciences*, *109*, 17490–17494.

Ross, L. N. (2015). Dynamical models and explanation in Neuroscience. *Philosophy of Science*, *82*, 32–54.

Ruiz-Mirazo, K., & Mavelli, F. (2007). Simulation Model for Functionalized vesicles: Lipid-peptide integration in minimal protocells. Advances in Artificial Life (pp. 32?41). Springer.

Sims, M. (2021). A continuum of intentionality: Linking the biogenic and anthropogenic approaches to cognition. *Biology and Philosophy*, *36*, 51.

Sims, M., & Kiverstein, J. (2022). Externalized memory in slime mould and the extended (non-neuronal) mind. *Cognitive Systems Research*, *73*, 26–35.

Smith-Ferguson, J., & Beekman, M. (2020). Who needs a brain? Slime moulds, behavioural ecology and minimal cognition. *Adaptive Behavior*, *28*, 465–478.

Suzuki, K., & Ikegami, T. (2009). Shapes and self-movement in protocell systems. *Artificial Life*, *15*, 59–70.

Van Duijn, M., Keijzer, F., & Franken, D. (2006). Principles of minimal cognition: Casting cognition as sensorimotor coordination. *Adaptive Behavior*, *14*, 157–170.

van Eck, D. (2018). Rethinking the explanatory power of dynamical models in cognitive science. *Philosophical Psychology*, *31*(8), 1131–1161. https://doi.org/10.1080/09515089.2018.1480755.

Van Gelder, T., & Journal of Philosophy Inc. (1995). What might Cognition be. If Not Computation? *Journal of Philosophy*, *92*, 345–381.

Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Currents in Modern Biology*, *5*, 187–196.

Webre, D. J., Wolanin, P. M., & Stock, J. B. (2003). Bacterial chemotaxis. *Current Biology*, *13*, R47–R49.

Authors and Affiliations

Nick Brancazio[1] · **Russell Meyer**[2]

✉ Nick Brancazio
nick.brancazio@adelaide.edu.au

Russell Meyer
russell.meyer92@gmail.com

[1] University of Adelaide, Adelaide, SA 5005, Australia

[2] Chinese Academy of Sciences, Institute of Philosophy (CASIP), Beijing, China