

Minds, Machines, and Molecules

T. D. P. Brunet and Marta Halina
Department of History and Philosophy of Science
University of Cambridge

11 September 2020

Keywords Metabolism, Sentience, Molecular Machines, Synthetic Biology, AI, Functionalism

Abstract

Recent debates about the biological and evolutionary conditions for sentience have generated a renewed interest in fine-grained functionalism. According to one such account advanced by Peter Godfrey-Smith, sentience depends on the fine-grained activities characteristic of living organisms. Specifically, the scale, context and stochasticity of these fine-grained activities. One implication of this view is that contemporary artificial intelligence (AI) is a poor candidate for sentience. Insofar as current AI lacks the ability to engage in such living activities it will lack sentience, no matter what its coarse-grained functions. In this paper, we review the case for fine-grained functionalism and show that there are contemporary machines that fulfil the fine-grained functional criteria identified by Godfrey-Smith, and thus are candidates for sentience. Molecular machines such as Brownian computers are analogous to metabolic activity in their scale, context and stochasticity, and can serve as the basis of AI. Molecular computation is a promising candidate for artificial sentience according to contemporary philosophical accounts of sentience.

1. Introduction

In an address to the members of the European Parliament, the philosopher Thomas Metzinger asks the EU to “ban all research that risks or directly aims at the creation of synthetic phenomenology” (Metzinger, 2018, p. 2). Metzinger argues that current artificial intelligence (AI) lacks political and ethical representation; thus, were researchers to create an artificial system capable of subjective experience such as suffering, we would lack the tools to mitigate the associated risks. Although Metzinger is not alone in his concerns regarding the creation of synthetic phenomenology, others believe artificial sentience is beyond our technological capacities (see Dennett, 1994, and Shanahan, 2015, for discussion).

One of the difficulties with these disagreements is that we lack a widely accepted account of sentience that we can apply to AI. Although such accounts are lacking, a recent set of approaches aims to gain a deeper understanding of the biological and evolutionary conditions for sentience (Feinberg & Mallatt, 2016; Jablonka & Ginsburg, 2019). According to these approaches, we can get a handle on the properties and functions of sentience by looking at its evolution and function in organisms. We use the terms ‘sentience’ and ‘subjective experience’ synonymously here to mean that there is something it feels like to be a system.¹ One advantage of this biological approach is that it supports understanding sentience across the animal kingdom. An account of sentience grounded in evolutionary history can be used to assess organisms that diverge widely in their structure and behaviour, such as honey bees, octopuses and great apes. In contrast, an approach that narrowly seeks to identify the physical correlates or markers of consciousness in humans alone may be of little use when it comes to understanding sentience in our more distant relatives, creatures that diverged from us hundreds of millions of years ago, such as the octopus (Godfrey-Smith, 2016).

We think an approach grounded in the evolution of life is broadly correct. However, we are also concerned that it could lead to mistaken conclusions regarding the possibility of artificial sentience. According to Peter Godfrey-Smith’s account (Godfrey-Smith, 2016a, 2016b, 2017, 2019), which we focus on here, sentience is realised by “living activity” and will not likely be realised in contemporary metal-and-silicon AI because such systems lack the relevant properties of living systems. We introduce Godfrey-Smith’s view in more detail in section 3, however, it would be worth briefly outlining what he takes to be the relevant properties of living activity:

1. Scale – living activity occurs at the cellular level
2. Context – living activity occurs dissolved in an aqueous medium
3. Stochasticity – living activity is in part stochastic or subject to Brownian motion

Godfrey-Smith writes, “If this [living activity] can be realized artificially, it would be achieved on a different path from that pursued in familiar AI and robotics projects” (2016a, p. 505). In what follows, we challenge this claim by showing that current machines fulfil the fine-grained functional criteria identified by Godfrey-Smith, and thus are candidates for sentience.

¹ Sentience in this sense has something to do both with having a "perspective" or "point of view" and with having "feelings" or "sensations". We follow Godfrey-Smith (2016a) in using the term "consciousness" to refer to a more complete or fully fledged experiential capacity. That said, these terms are difficult to define in ways broadly satisfying to both philosophers and evolutionary biologists, especially in the very diverse contexts of widely divergent species with natural minds and hypothetical artificial ones. We do not think our arguments turn on the particular definition of subjective experience used here.

One of the reasons it might be difficult to imagine AI exhibiting properties characteristic of living activity is that discussions of AI tend to focus on a small group of systems: familiar machines such as clockworks and ordinary circuitry. The remedy for this slender diet of examples is to consider a wider domain of machines. Computation has been embodied in a diversity of machines from self-healing gelatinous sheets (Urban, 2012) to biopolymers in test tubes (Wang et al., 2006; Landweber, 1998) and self-organizing and adaptive nanoscale oscillators (Yogendra et al. 2017; DARPA 2016).² Once we widen our set of examples to include some of these unfamiliar machines and alternative instantiations of computation, it becomes far less clear what feature could be shared by all biological minds to the exclusion of artificial ones.

This paper proceeds as follows: In section 2 we briefly review the import of functionalism in debates about the capacity for artificial minds, concentrating on the shift from coarse-grained to fine-grained functionalism. In section 3 we characterize the view advanced by Godfrey-Smith, which combines fine-grained functionalism with an evolutionary approach to understanding the conditions for sentience. There we identify three relevant properties of living activity that make living systems candidates for sentience, according to Godfrey-Smith: scale, context, and stochasticity. In section 4 we show that there are artificial systems that satisfy these properties. Here we draw on examples of molecular machines, such as biomolecular ratchets, rotaxanes, catenanes, and so-called DNA-origami, as well as computational processes that are realized in aqueous, DNA-based, stochastic and Brownian systems. We argue that these machines are candidates for artificial sentience. In section 5 we reply to two objections to the claim that current machines are candidates for sentience: first, that artificial systems do not evolve, and second, that artificial systems fail to engage in some other activity characteristic of living systems. We argue that neither of these objections succeeds. Thus, if sentience depends on living activity, then artificial sentience may very well arise in contemporary AI and robotics research.

2. Matter and Functionalism

Recent advances in AI have resulted in computer programs with surprisingly sophisticated behaviours, such the ability to drive a car or play chess and Go at superhuman levels (Badue et al., 2019; Silver et al., 2018). According to coarse-grained functionalism, programs such as these have mental capacities similar to humans insofar as they have a similar coarse-grained economy of inputs, outputs and internal states (see

² DARPA's Real Time Machine Learning (RTML) and Unconventional Processing of Signals for Intelligent Data Exploitation (UPSIDE) initiatives have reported a number of unconventional forms of computation (obtained from: <https://www.darpa.mil> 22/04/2020).

Sprevak 2009). These functions are “coarse-grained” in the sense that they abstract away from various details that might differ across individuals. Two systems can be materially different while performing the same coarse-grained function, such as when a book or a magnet serves as a doorstop. John Searle advanced his now famous argument against such a functionalist view in 1980. The problem with coarse-grained functionalism, he argued, is that it is too liberal in its attribution of mental states. Under such a view, as Hilary Putnam noted, “we could be made of Swiss cheese and it wouldn’t matter” (Putnam, 1975, p. 291).

Although Searle objected to coarse-grained functionalism, he did not deny the possibility of thinking machines. Quite the contrary, he wrote, “*only* a machine could think, and only very special kinds of machines, namely brains and machines with internal causal powers equivalent to those of brains” (Searle, 1980, p. 417). When responding to the question of whether a machine engineered by humans could think, Searle said yes, “assuming it is possible to produce artificially a machine with a nervous system...sufficiently like ours” (Searle, 1980, p. 422). Between “equivalent to” and “sufficiently like” there is a world of difference. The carbon based biochemistry of humans is in certain ways similar to that of silicon, and human metabolisms are in certain ways similar to the electro-chemistry of batteries, but these similarities do not imply that silicon chips can think any more than they imply that we can power a remote.³ Simply asserting that the causal powers or biochemical nature of brains are relevant to our explanation of subjective experience does not, by itself, imply that ordinary machines are not similar to brains in precisely the relevant ways. We must also say what those ways are: what is relevant about being a system built of this or that material.

Like Searle, Godfrey-Smith rejects coarse-grained functionalism for understanding the mind and sentience. Unlike Searle, however, Godfrey-Smith provides an account of those fine-grained functions that might be important for sentience. Materials also matter for Godfrey-Smith insofar as they can or cannot perform the necessary fine-grained functions. Fine-grained functions impose constraints on matter because different materials have different affordances and behavioural capacities. Water has sufficient electrostatic interactions to dissolve certain carbon molecules; silicon dioxide will not have the same effect, since it does not have the same properties.

As the battery example above illustrates, humans share all sorts of fine-grained functions with other objects in the world. Showing that a system, such as an octopus, shares some

³ Human metabolism and batteries both require the transfer of electrons and elements of acid-base chemistry. Incidentally, most organisms can be used to extract an electric current (like a battery) under the right conditions of oxidation, but the point is that this sort of happenstance similarity is not, in itself, relevant to the explanation of why we have mental states.

fine-grained functions with humans is insufficient to demonstrate that this system is sentient. Likewise, showing that a system, such as a computer, *differs* in its fine-grained functions to humans is insufficient to establish its lack of sentience. One must instead make a case regarding which fine-grained functions are relevant to sentience such that the presence or absence of these functions tracks the presence and absence of sentience. Luckily Godfrey-Smith provides us with such an account. In the following section, we introduce this account before applying it to artificial systems.

3. Living Activity

Godfrey-Smith argues that sentience is bound up with the fine-grained functions characteristic of life. He writes: “The argument made here is not that the particular chemical elements and molecules must matter. They might be hard to replace in some cases, easier in others. But the functional profile that would have to be realized includes living activity” (Godfrey-Smith, 2016a, p. 505). According to Godfrey-Smith, living systems have a unique suite of fine-grained functions or activities that are important for sentience. Insofar as AI lacks the ability to engage in such living activities it will lack sentience, no matter what its coarse-grained functions. After introducing the functions that Godfrey-Smith thinks are important for sentience in this section, we argue in the following section that these functions are not unique to living systems, but are found in contemporary AI.

Godfrey-Smith argues for a chain of dependencies linking physical matter to metabolic activity and metabolic activity to sentience. Starting with physical matter, there are material features that are critical for metabolic success, such as a particular scale (nanometers) and context (solvent). This scale and context allows matter to behave “differently from how it behaves elsewhere” (Godfrey-Smith, 2016a, p. 485). In this nanoscale aqueous world, there is spontaneous motion, vibrating molecules, attractions, repulsions, diffusion, stochastic forces, and energetic effects not found in the world of middle-sized dry goods. The result is a “molecular storm” in which “the way things get done is by biasing tendencies in the storm” (Godfrey-Smith, 2016a, p. 485). In section 4 we refer to this idea of “biasing tendencies in the storm” as “ratcheting”.

Godfrey-Smith argues that such nanoscale, dissolved, and stochastic systems are different from familiar machines like clocks and computers (see also Skillings, 2015). Further, that it would be difficult, if not practically impossible, to build a metabolic system like our own outside of this scale and context. In other words, properties of scale, context and stochasticity are important for the emergence and subsistence of metabolisms in biological organisms. Metabolisms in turn sustain other important fine-grained functions, such as internally coordinating the parts and activities of a system, utilising materials and

energy to maintain a system's organisation, and sensing and responding to appropriate environmental stimuli. Functions such as these enable an organism to regenerate, maintain its boundary, respond flexibly to disruptive events, and other activities needed to stay alive. A "minimal subjectivity" or "point of view" emerges from these activities as the organism works to ensure it remains distinct from its environment (Godfrey-Smith, 2016a, p. 491).

According to Godfrey-Smith, minimal subjectivity is a step in the direction of sentience, but not yet sufficient for it. Sentience or subjective experience requires "sensing and action of a richer kind" (Godfrey-Smith, 2016a, p. 495). Living organisms must coordinate their activities with a wide range of things in order to remain alive. They must cope with changing conditions, both internal and external, in a world in which the processes that sustain them affect these conditions. Many organisms cope with this changing landscape by using their senses and nervous systems: they seek prey, avoid predators, and detect internal states such as tissue damage, hunger and thirst. Coordinating with cues in this way involves cognitive activity that likely feels like something from the inside, perhaps resembling something like having a point of view and distinguishing good from bad (Godfrey-Smith, 2019, p. 23).

Much of this is so even in organisms without large brains and central nervous systems. Bacteria, such as *E. coli*, engage in simple forms of coordinating behaviour. For example, the "tumble and run" strategy involves sensing the ambient nutrient content and engaging in one of three actions in response: running in a straight line, "tumbling" around randomly and then running, or staying put (Bechtel, 2014; Godfrey-Smith, 2017). A bacterium retains information about the ambient nutrient content before its last run and can use this along with information about its current state to move in the direction of higher nutrient content. Godfrey-Smith refers to capacities such as these as "proto-cognitive" (Godfrey-Smith, 2016a, p. 490). They represent the evolutionary stage before sophisticated sensorimotor capabilities emerged. Sophisticated sensorimotor capacities typically involve multicellularity, complex sensory organs, body parts, nervous systems and other structures that enable organisms to speed up and refine their internal and external communication and coordination. With these latter capacities likely come richer forms of sentience (or perhaps sentience for the first time), but they are woven from the same fabric as proto-cognitive abilities: metabolic storms with the capacity to self-organise and self-produce, harnessing materials and energy to keep a system alive.

The above provides a sketch of the main features of Godfrey-Smith's account (see Godfrey-Smith 2016a, 2016c and 2019 for more details). It is worth pausing here to highlight the nature of this account. Godfrey-Smith is not supplying necessary and

sufficient conditions for sentience. Instead, he is advancing what he calls a “how-possibly-necessarily explanation”. He writes:

Especially with respect to its details, the evolutionary sketch in this section should be seen as a how-possibly explanation. But it's a little more than that; the aim has been to walk through a series of steps such that once we reach one stage, we can see the next ones waiting. The result could then be called a how-possibly-necessarily explanation. Here the first modality – “possibly” – is epistemic. The second modality is not a strong sense of necessity, but something that contrasts with mere accident – a kind of robustness or predictability. It's a story about how things might have trodden a predictable path, on a planet like ours, a path by which subjectivity reliably arose. (Godfrey-Smith, 2019, p. 26)

Godfrey-Smith's account shows not only how sentience *could* have arisen in organisms like us, but the conditions in place which make this a robust, predictable or non-accidental path to sentience.

One consequence of the above account, according to Godfrey-Smith, is that current AI and robotic systems are unlikely candidates for sentience. One can see why: the materials from which familiar AIs are constructed do not engage in the right fine-grained functions. However, we think the Wittgensteinian admonition not to consider too one-sided a diet of examples applies here (Wittgenstein, 1953/2009, p. 164). The class of machines that might serve as a material basis for AI is wider than commonly supposed. In the following section, we introduce such a broader class of machines--a class that includes strong candidates for sentience according to the account advanced by Godfrey-Smith.

Before closing this section, it is worth emphasising the empirical and historically contingent (Rapp, 2012), rather than *a priori*, nature of the question: “are current AI and robotic systems likely candidates for sentience?” There is nothing about the concept of machines that excludes them *a priori* as candidates for sentience, although it might of course be an empirical matter of fact that contemporary AI lacks the features relevant to sentience. Ian Hacking (1998), writing about Georges Canguilhem, put this well. After noting that scholars like Descartes hold that animals cannot engage in certain acts in principle (like thinking), Hacking writes:

For Canguilhem these are not the fundamental distinctions. Even questions such as ‘Can machines think?’ might be answered, in the spirit of his writing, ‘Well, have we in fact made any thinking machines yet?’ We have to ponder the matter, but we should not start from a fundamental opposition between machine and organism. (Hacking, 1998, p. 208)

Godfrey-Smith is in good company with Canguilhem and Hacking. At no point does he claim that sentience cannot be realized artificially, just that it must be realized on a “different path” from familiar AI. Such a different path must include not only the right coarse-grained functions for mindedness (such as perhaps having goals and perceptions), but also the right fine-grained ones (the right scale, context and level of stochasticity). Living activities are important, not because they are infused with “*élan vital*” or other mysterious qualities, but because they are unique in the functions they can perform. There is no “fundamental opposition between machine and organisms” for Godfrey-Smith. Instead, machines and organisms differ empirically in ways that happen to matter for sentience.

We can express the idea that there is no fundamental or *a priori* opposition between machines and living organisms in the form of a fair-treatment principle.⁴

Fair-treatment principle: If an artificial process is relevantly similar to a sentient biological process (save for being artificial), then that artificial process should have an equal claim to sentience.

We part ways with Godfrey-Smith in holding that current machines and biological systems are similar in precisely those ways that Godfrey-Smith identifies as relevant for sentience. The machines with the relevant fine-grained functions are broadly known as “molecular machines.” These machines are constructed on a nanoscale and behave stochastically in aqueous and similar environments. According to Godfrey-Smith’s account of sentience and the fair-treatment principle, we should conclude that AI built from molecular machines have an equal claim to sentience.

4. Molecular Machines

In 1959, Richard Feynman gave a lecture to the American Physical Society titled, “There’s Plenty of Room at the Bottom” (Feynman, 1960). The talk implored researchers to explore the world of the very small. Feynman argued that the nanoscale world would contain many hitherto unexplored phenomena and engineering opportunities, as did the field of low-temperature physics. Inspired by the “marvellous biological system”, Feynman wrote that “biology is not simply writing information; it is *doing something* about it... [cells] are very active; they manufacture various substances; they walk around; they wiggle; and do all kinds of marvellous things—all on a very small scale” (p. 25). Synthetic chemists, biologists and engineers today cite Feynman’s lecture as inspiring the

⁴ This principle follows Mark Sprevak (2009) who advances a fair-treatment principle for evaluating extended cognition (following Clark and Chalmers’s “parity principle”). We have substituted ‘internal’, ‘external’ and ‘cognitive’ for ‘biological’, ‘artificial’, and ‘sentient’ respectively (Sprevak, 2009, p. 3).

now active and growing field of molecular computing and nanotechnology (Sluysmans and Stoddart, 2018).

In this section, we show how machines from this field exhibit fine-grained functions thought to be characteristic of metabolism in sentient organisms. We consider three forms of molecular machines below: mechanically interlocked molecular architectures (MIMAs), aqueous computing, and molecular ratchets. As we progress through these examples we note how they successively possess more of these relevant properties, so that MIMAs have the relevant scale, aqueous computers have the relevant scale and context, and finally molecular ratchets exhibit the relevant scale, context and level of stochasticity. We hope these examples will not only expand standard conceptions of AI as metal-and-silicon machines, but also render plausible the idea of artificial sentience, given Godfrey-Smith's account of sentience.

4.1 Scale: Mechanically Interlocked Molecular Architectures

Mechanically Interlocked Molecular Architectures (MIMAs) are a class of large molecular complexes wherein molecules that are not necessarily (covalently) bonded to each other are nonetheless linked due to their topology. Examples include *catenanes* and *rotaxanes*, as well as molecular analogues of *knots* and *rings*. A catenane is any molecule wherein cyclic chains are linked three dimensionally such as a pair of linked rings. A rotaxane is any molecule with a “ring” and “axel”, i.e. where a closed circular molecule (the ring) wraps a chain of another molecule (the axel) such that it can rotate but will not “fall” off the chain.

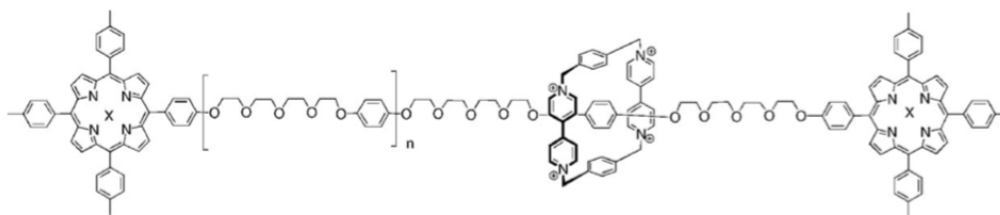


Figure 1. Rotaxane. The long molecule spanning horizontally is the “axel”; the molecule slightly offset to the right is the “ring”. “Caps” or “dumbbells” on each end prevent the ring from falling off the axel. Image reproduced from Fiaz et al. (2009).

MIMAs have applications ranging from fluorescent tagging, contracting molecular “muscles”, and nanoscale “cars” with directional control to logic-gates and molecular data storage devices (Chen et al., 2003; Green et al., 2007; Ma and Tian, 2010; Bruns and Stoddart, 2014). In the latter case, chemical, electromagnetic or optical energy can be used to change the configuration of interlocked molecules, storing information. Indeed, MIMAs like rotaxanes have been heralded as potentially the most dense data-storage

medium possible, outstripping conventional memory circuitry by orders of magnitude (Cox 2001).

These components of molecular machines show that function at the nanoscale is not specific to the parts of organisms. With more complex arrangements of MIMAs, the circuitry and circuit-like systems used in AI research can be constructed at the molecular scale. Insofar as sentience requires this scale of living activity, this places molecular machines in a category potentially ripe for proto-cognition and subjective experience, provided they can also operate in a similar context and exhibit a similar level of stochasticity to living systems. Examples of such latter systems come from the field of aqueous and DNA computing, discussed next.

4.2 Scale and Context: Aqueous Computing

Aqueous computing is computation that occurs in water (Head et al., 2002). Two major advantages of this form of computing are its information density and economy of energy (Gibbons et al., 1997). One of the most successful forms of aqueous computing used thus far is DNA computing. DNA computing was conceived by the physicist Mikhail Samoilovich Neiman in his 1960s discussion of polymeric “molecular memory” systems (Brunet, 2016). Thirty years later, the computer scientist Leonard Adleman pioneered experimental DNA computing by showing that he could use it to solve a version of the “traveling salesman problem” (Adleman, 1994). Adleman’s experiments used DNA assembly in vitro to compute the Hamiltonian path of a directed graph—i.e., the path in which each node of the graph is visited exactly once. The power of DNA computing to engage in massively parallel searches with much greater energy efficiency than a supercomputer became apparent thereafter (Adleman, 1994; Gibbons et al., 1997).

As a polymer, DNA has a number of properties that make it a good candidate for molecular computing experiments. Its manipulation, synthesis and sequencing is well understood from past molecular biological research; it can store arbitrary information appropriately encoded in its sequence (Church et al., 2012; Goldman et al., 2013); it can be organised to create fast logic gates and signal transmission lines (Chatterjee et al., 2017); and it can fold into an immense variety of shapes which can themselves bind to neighbouring folded molecules to give rise to even more complex gates (Sha et al., 2005; Seeman, 2007).

Folded DNA structures, sometimes called ‘DNA origami’ (Castro et al., 2011), can take on a rich variety of forms, from images of smiling faces or dolphins to flexible machine-like components and knots. In these systems, folding and computation are intimately related. One way to perform computation with DNA is to synthesize a sequence that folds into the three dimensional structure of Wang tiles (Wang, 1961; Woods et al., 2019).

These are a formal system for implementing algorithms where computational processes are converted into problems of matching adjacent faces of a pre-defined and pre-synthesized set of “tiles”, converting computation into molecular origami. Moreover, at sufficient levels of complexity, some such DNA systems are provably Turing complete (Varghese et al., 2015).

These sorts of chemical computational processes do not need to be based in DNA. What is essential to this sort of computation is that one uses a polymer that can engage in sequence-dependent binding to neighbours, which has been achieved with nucleic acid analogues and other more exotic chemicals. For instance, Yamamura et al. (2001) report a method of aqueous computation using peptidyl nucleic acid (PNA)—an analogue of DNA where the usual backbone of DNA is replaced by linked peptides. Regardless of the particular molecular constitution, whether derived from biomolecules or synthesized *ab initio*, these alternative varieties of computation are embodied at the nanoscale and in contexts where components are variously dissolved. As Feynman noted in his lecture, matter behaves differently at this scale and context; engineers have had remarkable success harnessing these behaviours to solve problems in new and efficient ways.

4.3 Scale, Context and Stochasticity: Molecular Ratchets

For any system that must do work under the influence of chaos there are two options: find a way to mediate the effects of this influence, or find a way to turn chaos to your service. Evolution is remarkably good at producing molecular structures deploying the latter strategy, that function by “biasing tendencies” of the molecular storm inside cells at physiological conditions. Molecular ratchets are one important way this is achieved (Oster, 2002; Ait-Haddou and Herzog, 2003). Like the common tool of the same name, ‘ratcheting’ is a general process where the effects or outputs of a system are directional despite the fact that the cause or input is not. Importantly, ratchets do not function *despite* non-directional inputs, by mediating or reducing them, but by having particular causal processes that take advantage of them.

A prime example of a molecular ratchet is ATP-synthase, a protein complex at the heart of metabolism. In this system a channel between membrane compartments allows the diffusion of hydrogen ions, which is a stochastic or Brownian process. This passage of ions tends to result in a circular rotation of the F_0 component of ATP-synthase and drives downstream production of ATP from ADP (Nakamoto et al., 2008; see Wideman et al., 2019). Ratcheting is a way of generating biased tendencies from an unbiased or chaotic external influence, but it is not specific to biomolecules. At the macro-level we can see similar energy generating ratchet processes in self-winding watches, where the ambient movement of the wearer’s body causes an oscillating weight or rotor to swing. These

chaotic movements wind a coiled strip of metal (the mainspring), which then has directional (clockwise) downstream effects, providing the watch with a constant supply of mechanical energy. Ratcheting is an essential process guiding the fine-grained stochastic effects of biomolecules, and on Godfrey-Smith's view is important to subjective experience of the sort enjoyed by organisms (see section 3), but it is also a feature of the functional profile of many machines.

At the macro-level again, artificial neural networks (ANNs) are digital ratcheting devices in many ways analogous to biological molecular ratchets. Godfrey-Smith notes that biological causal processes “are based in networks with many redundancies and small effects. These have consequences for robustness and adaptability” (Godfrey-Smith, 2016a, 503). This is also true for ANNs. They can be fed large quantities of fairly disorganized information, which is passed along the network while being “nudged” by weights in directions deemed useful during the largely statistical processes of training. This creates a biased tendency to produce a certain kind of output, despite consisting of a large number of meaningless or dispensable processes.

It is somewhat unsurprising that ANNs work this way. Both learning and evolutionary processes tend to result in satisfying or “good enough” solutions to complex problems and, in complex systems especially, biasing a tendency is often an easier solution than overcoming or eliminating the chaos involved in existing processes. In the words of the biologist François Jacob (1977), evolution is a tinkerer; it produces novelty by constrained alteration of existing structures. The statistical processes involved in training ANNs are sophisticated tinkering, methodically adjusting parameters until the computational process outputs a desired result to a satisfying degree of accuracy.

ANNs are particularly good examples of ratcheting processes that enable computation despite stochastic inputs, since they are also *prima facie* candidates for artificially intelligent and sentient systems, provided Godfrey-Smith is correct about the importance of scale, context and stochasticity. Nonetheless, the need to manage stochastic processes is not specific to ANNs and occurs whenever computation takes place. Godfrey-Smith acknowledges that low-level changes in computers are inevitable, but these are “engineered to be as small as possible” and this makes computers “different from living systems in ways that make engineering sense” (Godfrey-Smith, 2016a, p. 503). However, this is not always the route taken in computer engineering, nor always the best method for achieving effective computation, as the examples in this section show.

Instead of being limited or engineered away, chaotic effects are regularly exploited in the field of stochastic computing. For von Neuman (1956), the important point of stochastic computation is the same as for probabilistic logic: that unavoidable error in parts must

somehow be accounted for in the production of useful wholes. Indeed, von Neumann is clear that this stochasticity is also a non-accidental part of advanced computer automata, which he and many in his cohort fondly referred to as “organisms”. He writes that error in computer automata is seen, “not as an extraneous and misdirecting accident, but as an essential part of the process under consideration” (von Neumann, 1956, p.43). No computational process—digital, DNA or quantum—produces zero entropy. All entail the production of at least some heat or delay, and the consequent potential for “error” in computation. A large part of computer science deals with how to cope with computational errors due to lower-level stochastic processes. Though the aim—identified by von Neuman—is not simply to limit low-level changes but to recognize and exploit their fundamentality to the system.

As a final example, Brownian computation methods in theoretical computer science have taken the route of turning chaos into an advantage. As the physicist and information theorist Charles H. Bennett writes:

In these [Brownian] computers, a simple assemblage of simple parts determines a low-energy labyrinth isomorphic to the desired computation, through which the system executes its random walk, with a slight drift velocity due to a weak driving force in the direction of forward computation [p. 905]...the Brownian computer makes logical state transitions only as the accidental result of the random thermal jiggling of its information-bearing parts. (Bennett, 1982, p. 912)

This description of Brownian computation sounds remarkably like Godfrey-Smith’s characterisation of the ratcheting, biased tendencies of metabolism. In order to distinguish the operations of fundamental metabolic components from familiar machines, Godfrey-Smith says that the processes typical of such biomolecules, a “storm-like collection of random walks influenced by friction, charge, and thermal effects” are “non-mechanistic” (Godfrey-Smith, 2016a, p.486). However, it is precisely those features of biomolecules that Godfrey-Smith identifies as “non-mechanistic” and essential to the operation of metabolic processes, that are essential to the operation of machines like the Brownian computer. Like Godfrey-Smith, Bennett also notes that, however unlikely such processes are in the macroscopic world, this way of getting things done is common at the level of molecular reactions (Bennett, 1982, p. 912).

More recent work on Brownian computation has continued to draw inspiration and theoretical principles from Bennett and ratcheting in biomolecular motors. Peper et al. (2013) note that “organisms have mechanisms in place, not only to cope with fluctuations, but also to exploit them as a driving force in biological processes” (p. 2), which is the aim of Brownian computation. Peper et al. (2013) and Lee et al. (2016) focus

on designing universal Brownian circuitry capable of extracting useful computation from nano-scale fluctuations. These fluctuations are “not just a nuisance” (Lee et al. 2016 p. 342) but “actively exploited” [ibid]. Moreover, to make sure that computation proceeds at an appreciable rate despite inevitable backward processes, speed is increased by “ratcheting devices” that are “well-known in nature” (Peper et al. 2013, p. 19) and work by biasing probabilistic circuit transitions. Importantly, like many biomolecular ratchets, Brownian circuitry is unable to achieve certain effects *unless* there is sufficient stochasticity in the system.

Complex bio-derivative forms of Brownian computation are also exemplified in DNA computing. What remains to be seen is the achievement of full-scale AI in these alternative mediums. Qian et al. (2011), and later Cherry and Qian (2018), have taken some of the first proof of concept steps: they fabricated neural networks using DNA-based computational techniques. As Qian et al. (2011) point out, “Stochastic simulations suggest that the four-neuron DNA associative memory would function reliably [in] a volume of roughly $1\mu\text{m}^3$, that is, small enough to fit inside a bacterium” (p. 372). Here is a system that functions stochastically, dissolved in water, at the scale of nanometers, and at the same time has the potential to realize the coarse-grained computational “information processing” operations characteristic of traditional AI projects.

We have reason to believe this sort of AI research will continue; the practical benefits of such technology have been steadily accruing. Chen et al. (2015) review a number of biomedical applications of “dynamic” DNA nanotechnologies, i.e. those, like circuits, with time-dependent and analytic functionality. Moreover, the prospects raised by some of these technologies—that they might overcome limitations on size and energy-efficiency that currently limit familiar circuitry (Livstone et al. 2003)—suggest that their realization may be driven by demands within AI and robotics. This makes it more likely that these unfamiliar forms of AI will be achieved on a not-too-distant path from current AI projects.

We can look to these unconventional but well-theorized cases of machines for precisely the sorts of fine-grained functions thought relevant for the achievement of sentience. Perhaps our 18th or 19th century complement of machines was impoverished in functionality, but modern examples surpass these in both degree and kind. A DNA-scale computer stochastically checking multiple solutions to a graph-theoretic problem with dissolved components in parallel is about as far from a microprocessor as the latter is from the macroscopic ‘difference engines’ of Babbage or Lovelace. What fine-grained functional or material bases might be lacking in clockwork or contemporary metal-and-silicon computers can be found in these alternative embodiments of artificial intelligence.

If Godfrey-Smith's view that the metabolic side of life is relevant for sentience is correct, then molecular machines represent promising candidates for artificial sentience. In some cases, these machines exhibit all of the fine-grained functional properties--scale, context, and stochasticity--that Godfrey-Smith identifies as relevant for the emergence of sentience in living systems. We hope the examples discussed here help demonstrate that the answer to the empirical question, "are current AI and robotic systems likely candidates for sentience?" might be "yes."

5. Objections from Evolution and Life

5.1 Objection from Evolution

One major objection to the view that molecular machines are compelling candidates for sentience is that they lack a critical feature characteristic of living systems: biological organisms depend not only on metabolisms for their continued existence, but also on evolution for their origin. As Godfrey-Smith notes: "Life has a metabolic side and a side that has to do with reproduction and evolution" (Godfrey-Smith, 2016a 484). The molecular machines described above share features with metabolic systems, but are largely products of human engineering, not biological evolution. One might thus argue that, without this evolutionary component, artificial systems are not serious candidates for sentience.

We see three ways of proposing this objection. First, perhaps evolution is necessary for sentience. Second, perhaps evolution is required for sentience in some sense that is strong but nonetheless weaker than nomic or logical necessity; perhaps it is extremely *unlikely* for humans to engineer systems with functions analogous to those that, in organisms, required billions of years of evolution to arise. Third, one might argue that molecular machines lack the foundational materials for evolving sentience, that even if they were subjected to evolutionary forces, they would lack what is needed to evolve proto-cognition or minimal subjectivity.

In response to the first two arguments, some molecular machines indeed were and are subjected to evolutionary processes. The molecular machines designed by synthetic biologists have a mixed history, like dairy cows, owing their existence in part to artificial and in part to natural forces (Lewens 2013). If merely having some evolutionary component is required then DNA computers are safe from objections on the basis of past evolution. DNA itself and the enzymes used in computational protocols are, in part, outcomes of evolution.

If it is rather present evolvability that is required, we can look to the algorithms and experimental processes analogous to evolution that engineers use, for example, for solving optimization problems and discovering new molecular structures (Hu & Banzhaf, 2010). In a recent study, researchers used a genetic algorithm, machine learning, and in vitro analysis to explore the space of possible antimicrobial peptides or AMPs (Yoshida et al., 2018). AMPs are short (10-50 amino acids) peptides found in almost all living organisms, which kill microbes by breaking bacterial membranes, inhibiting DNA, RNA, and protein synthesis, and other methods. Researchers have long been exploring ways of designing new peptides for clinical use (Loose et al., 2006). The technique developed by Yoshida and colleagues shows how one can artificially evolve functional biomolecules by randomly generating a population of peptides and evaluating the “fitness” of individuals in silico and in vitro. A subset of this population is then selected, forming the basis of a new population in which mutations and crossovers are introduced (Yoshida et al., 2018).

Although the rates of biological and artificial evolution are difficult to compare, much work is dedicated to increasing the speed of artificial evolution for the benefit of practical applications (Hu & Banzhaf, 2010). Moreover, artificial evolution can be combined with methods like deep learning and reinforcement learning to efficiently develop new systems with surprising capacities. The program AlphaStar—the first system to defeat top human professionals at the real-time strategy game StarCraft—was developed in this way (Arulkumaran et al. 2019). If sentience depends on cumulative evolution (in either the stronger or weaker senses above), this is unlikely to place it beyond the reach of our current engineering efforts.

Regarding the third argument, we think it is important to be clear about the difference between gradualist and saltationist accounts of the emergence of fully-fledged consciousness. Our response is inspired by Samuel Butler, a contemporary of Charles Darwin. Butler proposes the descent of conscious machines as a disjunction elimination from the assumption that consciousness either emerges gradually *or* saltationally.

If consciousness is seen as saltational—occurring in a leap or “definite step”—then we are left with the mystery of explaining that leap. However, we know that such a leap is possible. As Butler put it, if consciousness emerges by a leap, then “the race of man has descended from things which had no consciousness at all” (Butler, 1872, p. 197). That is, if consciousness evolved saltationally, then something need not have any consciousness at all for it to evolve consciousness. In this case, current molecular machines might be strong candidates for consciousness without themselves having any consciousness at all. At least, so far as we know, given that we have no satisfying explanation for that leap in either machines or organisms.

If instead consciousness emerges gradually, Butler proposes, then much of the “action that has been called purely mechanical and unconscious”, such as that of our distant and not obviously conscious ancestors, must have contained “elements of consciousness” (Butler, 1872, p. 197). If so, we are left without a reason that other “purely mechanical” actions might not have elements of consciousness. Butler writes: “There is no security... against the ultimate development of mechanical consciousness, in the fact of machines possessing little consciousness now. A mollusk has not much consciousness” (Butler, 1872, p. 194). Godfrey-Smith and Butler share the position that consciousness can be found in diverse entities and that it either evolves gradually or exists widely *in embryo* or proto-cognitively. We imagine Godfrey-Smith largely agreeing with Butler about molluscs and other metabolizing cases, while denying that contemporary machines possess even a “little consciousness” now.

This claim about machines arises not because Godfrey-Smith denies the possibility of a system having a little consciousness. Godfrey-Smith believes one can have more or less consciousness and does not see the origin of consciousness as a “definite step” (Godfrey-Smith, 2016b, p. 68); he supports the view that subjective experience can come in “minimal but non-zero” proportions (Godfrey-Smith, 2016a, p. 495). Of course the notion of increments of consciousness or minimal degrees of subjectivity is not uncontentious (Shevlin forthcoming). However, gradualist evolutionary proposals of the sort endorsed by Godfrey-Smith and Butler require some such notion to get off the ground. On the assumption of gradualism, whether or not machines can evolve consciousness turns on whether they now possess the right foundations or building blocks for consciousness.

Butler’s analysis of consciousness in machines is motivated by epistemic considerations and concerns of risk. At a sufficiently proto-level, consciousness might be too subtle to notice, and thus to avoid or provide “security against” (Butler, 1872, p. 194). This leads him to attribute “germs of consciousness” to “many actions of the higher machines” (ibid., p. 197), as he would to molluscs. However, without some principled way of assessing “germs of consciousness” in both higher machines and molluscs, Butler provides little guidance for future analysis or standards for evaluating his own account.

Happily, Godfrey-Smith provides an account of what is required for minimal forms of subjectivity, and thus permits a deeper assessment of putatively similar candidates for the gradual evolution of consciousness. Godfrey-Smith is partial to *biopsychism*: the view that low-level conscious capacities are exhibited not by ordinary matter (classical panpsychism) or mechanical activities, but by “the simplest forms of cellular life” (Godfrey-Smith, 2016a, p. 495). Presumably, what is relevant about simple and

evolutionarily early forms of life are just those features of “living activity” thought relevant for sentience in organisms today. If we accept this framing, then the question of low-level conscious capacities in current machines becomes whether we can find “living activities” responsible for the “actions of higher machines”, as we would in ancient life, and our arguments in section 4 show that we can. If indeed these are the right fine-grained features of living activity required as a foundation or for the building blocks of minimal consciousness, then gradualist narratives for the origin of machine consciousness *can* get off the ground.

Butler, Godfrey-Smith, Canguilhem and Hacking would all agree that any inference to the descent of conscious machines from the present stock should be an *a posteriori* affair. Indeed, to Canguilhem’s or Hacking’s bare bones wait-and-see approach mentioned in section 3, Godfrey-Smith adds flesh in the form of the metabolic criteria required for living activity in simple cellular life.

5.2 *Objection from Life*

Finally, one might object to molecular machines as candidates for sentience on the grounds that they are not alive.⁵ Perhaps something other than the metabolic and evolutionary side of life is required for sentience--some property of living systems not highlighted in Godfrey-Smith’s account. This is a reasonable objection, given the nascent state of research on subjectivity. Godfrey-Smith weaves a rich tapestry of dependencies, linking the scale, context and stochasticity of molecular metabolisms to simple and complex forms of cognition. However, he is also quick to point out that the story provided is a preliminary sketch. Such a sketch requires revising and filling in, guided by our growing theoretical and empirical understanding of the relevant systems and processes. Given this, it is reasonable to expect new features relevant to sentience to emerge, and to reassess the question of artificial sentience in light of this research. We agree with this approach and offer our analysis of artificial sentience as one based on the current state of the art of research on the biological and evolutionary conditions for sentience.

We would like to emphasise, however, that if one objects to our claim that AI is a candidate for sentience on the basis that it lacks some other property of life, then one should also specify what this property is and why it is relevant. It will not suffice to say that these machines are not alive. The concept of life is ambiguous and can be filled out in different ways. One could thus appeal to this concept to get a desired outcome in an ad

⁵ Godfrey-Smith does not appeal to the concept of “life” in this way, instead specifying the activities and processes which he takes to be relevant to subjectivity. We think some readers, however, might be pulled in the direction of this objection, so address it here.

hoc manner. If one wanted to exclude AI from the class of sentient beings, for example, one could select precisely those living properties--particular aspects of growth or reproduction, for instance--that current AI lacks. Such an account would resemble what Deborah Mayo calls a “use-constructed” or “rigged” hypothesis or one constructed on the basis of known data, such as a marksman hypothesizing that she is a good shot by shooting holes in a board before proceeding to draw a target around those holes so that a bullseye is scored (Mayo 1996, p. 201). This method of theory construction should be avoided, as it cannot fail to fit the data, even if it were false.

Instead we think one should follow Godfrey-Smith’s lead and specify those fine-grained functions and broader evolutionary processes that we have reason to believe resulted in systems’ sentience. Once these are specified, they can be applied to other systems, as we have done here. We should however also keep in mind what Shevlin (forthcoming) refers to as the “specificity problem” or “the problem of how we can identify the appropriate level of detail that should be adopted by theories of consciousness in applying them to non-humans” (p. 1; see also Birch 2020). The more finely we spell out the criteria for sentience, the more committed we are to excluding systems that differ from those used to build our account (Sprevak, 2009). Explaining *why* a particular feature or function is relevant for sentience helps avoid this problem to some degree. In providing such explanations, one justifies the inclusion of a feature or process (at some given level of detail) in one’s account. Godfrey-Smith provides a compelling case that scale, context and stochasticity are relevant to the evolution of sentience. Any additional properties--whether features of life or not--should be similarly justified.

6. Conclusion

Concerns about the risks associated with artificial sentience call for the application of accounts of sentience to contemporary AI and robotics. Recent work on the origin of subjectivity in living organisms has produced evolutionarily and metabolically grounded theories of sentience. One such account, that advanced by Godfrey-Smith, suggests that familiar machines are unlikely candidates for sentience. In this paper, we have adopted Godfrey-Smith’s view that metabolic living activities--characterized by their unique scale, context and stochasticity--represent the right fine-grained functions for assessing sentience. On these grounds, however, we have further argued that molecular machines are candidates for sentience. Contemporary engineering of molecular machines has surpassed the material embodiments of earlier AI, giving rise to machines with the fine-grained features of living things.

We take the evolutionary and fine-grained analysis of sentience to be one of the best theories of the biological origin of sentience. Therefore, one of our best theories of

sentience implies that it might arise in molecular machines as well. Insofar as molecular machines with fine-grained functions characteristic of metabolic activity continue to be deployed in contemporary AI and robotics projects, artificial sentience or synthetic phenomenology may be achieved in the not too distant future.

References

- Adleman, L. M. (1994). Molecular computation of solutions to combinatorial problems. *Science*, 1021-1024.
- Ait-Haddou, R., & Herzog, W. (2003). Brownian ratchet models of molecular motors. *Cell Biochemistry and Biophysics*, 38(2), 191-213.
- Arulkumaran, K., Cully, A., & Togelius, J. (2019, July). Alphastar: An evolutionary computation perspective. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion* (pp. 314-315).
- Badue, C., Guidolini, R., Carneiro, R. V., Azevedo, P., Cardoso, V. B., Forechi, A., ... & Veronese, L. (2019). Self-driving cars: A survey. arXiv preprint arXiv:1901.04407.
- Bechtel, W. (2014). Cognitive biology: Surprising model organisms for cognitive science. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36), 158-163.
- Bennett, C. H. (1982). The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21(12), 905-940.
- Birch, J. (2020). The search for invertebrate consciousness. *Nous*, 1-21.
- Brunet, T. D. (2016). Aims and methods of biosteganography. *Journal of Biotechnology*, 226, 56-64.
- Bruns, C. J., & Stoddart, J. F. (2014). Rotaxane-based molecular muscles. *Accounts of Chemical Research*, 47(7), 2186-2199.
- Butler, S. (1863). Darwin among the machines. *ChristChurch PRESS, 13 June, reprinted by Mr. Festing Jones in, The Note-Books of Samuel Butler (Fifield, London, 1912, Kennerley, New York).*

- Butler, S. (1872) *Erewhon: Over the Range*. AC Fyfield revised edition. Gutenberg eBook (First published 1872)
- Castro, C. E., Kilchherr, F., Kim, D. N., Shiao, E. L., Wauer, T., Wortmann, P., ... & Dietz, H. (2011). A primer to scaffolded DNA origami. *Nature Methods*, *8*(3), 221.
- Chatterjee, G., Dalchau, N., Muscat, R. A., Phillips, A., & Seelig, G. (2017). A spatially localized architecture for fast and modular DNA computing. *Nature Nanotechnology*, *12*(9), 920.
- Chen, Y., Jung, G. Y., Ohlberg, D. A., Li, X., Stewart, D. R., Jeppesen, J. O., ... & Williams, R. S. (2003). Nanoscale molecular-switch crossbar circuits. *Nanotechnology*, *14*(4), 462.
- Chen, Y. J., Groves, B., Muscat, R. A., & Seelig, G. (2015). DNA nanotechnology from the test tube to the cell. *Nature Nanotechnology*, *10*(9), 748.
- Cherry, K. M., & Qian, L. (2018). Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks. *Nature*, *559*(7714), 370-376.
- Church, G. M., Gao, Y., Kosuri, S. (2012). Next-Generation Digital Information Storage in DNA. *Science*, *337* (6102): 1628. doi: 10.1126/science.1226355
- Cox, J. P. (2001). Long-term data storage in DNA. *TRENDS in Biotechnology*, *19*(7), 247-250.
- Dennett, D. C. (September 1-3, 1994). Consciousness in Human and Robot Minds. *IIAS Symposium on Cognition, Computation and Consciousness*.
- Feinberg, T. E., & Mallatt, J. M. (2016). *The ancient origins of consciousness: How the brain created experience*. MIT Press.
- Feynman, R. P. (2018). There's plenty of room at the bottom: An invitation to enter a new field of physics. In *Handbook of Nanoscience, Engineering, and Technology* (pp. 26-35). CRC Press.
- Gibbons, A., Amos, M., & Hodgson, D. (1997). DNA computing. *Current Opinion in Biotechnology*, *8*(1), 103-106.
- Godfrey-Smith, P. (2016a). Mind, matter, and metabolism. *The Journal of Philosophy*, *113*(10), 481-506.

Godfrey-Smith, P. (2016b). *Other Minds: The Octopus, The Sea, and The Deep Origins of Consciousness*. Farrar, Straus and Giroux.

Godfrey-Smith, P. (2016c). Animal evolution and the origins of experience, In D. Livingstone Smith (ed.), *How Biology Shapes Philosophy: New Foundations for Naturalism*. Cambridge: Cambridge University Press, pp. 51-71.

Godfrey-Smith, P. (2019). Evolving across the explanatory gap. *Philosophy, Theory, and Practice in Biology*, 11: 1-24.

Goldman, N., Bertone, P., Chen, S., Dessimoz, C., LeProust, E. M., Sipos, B., & Birney, E. (2013). Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*, 494(7435), 77.

Green, J. E., Choi, J. W., Boukai, A., Bunimovich, Y., Johnston-Halperin, E., DeIonno, E., ... & Tseng, H. R. (2007). A 160-kilobit molecular electronic memory patterned at 10 11 bits per square centimetre. *Nature*, 445(7126), 414.

Hacking, I. (1998). Canguilhem amid the cyborgs. *Economy and Society*, 27(2-3), 202-216.

Head, T., Chen, X., Yamamura, M., & Gal, S. (2002). Aqueous computing: a survey with an invitation to participate. *Journal of Computer Science and Technology*, 17(6), 672.

Jablonka, E., & Ginsburg, S. (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. MIT Press.

Jacob, F. (1977). Evolution and tinkering. *Science*, 196(4295), 1161-1166.

Landweber, L. F. (1998, May). The evolution of DNA computing: nature's solution to a path problem. In *Proceedings IEEE International Joint Symposia on Intelligence and Systems (Cat. No. 98EX174)* (pp. 133-139). IEEE.

Lee, J., Peper, F., Cotofana, S. D., Naruse, M., Ohtsu, M., Kawazoe, T., ... & Kubota, T. (2016). Brownian Circuits: Designs. *International Journal of Unconventional Computing*, 12.

- Lewens, T. (2013). From bricolage to BioBricks™: Synthetic biology and rational design. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 44(4), 641-648.
- Livstone, M. S., van Noort, D., & Landweber, L. F. (2003). Molecular computing revisited: a Moore's Law?. *TRENDS in Biotechnology*, 21(3), 98-101.
- Loose, C., Jensen, K., Rigoutsos, I., & Stephanopoulos, G. (2006). A linguistic model for the rational design of antimicrobial peptides. *Nature*, 443(7113), 867-869.
- Ma, X., & Tian, H. (2010). Bright functional rotaxanes. *Chemical Society Reviews*, 39(1), 70-80.
- Nakamoto, R. K., Scanlon, J. A. B., & Al-Shawi, M. K. (2008). The rotary mechanism of the ATP synthase. *Archives of Biochemistry and Biophysics*, 476(1), 43-50.
- Oster, G. (2002). Brownian ratchets: Darwin's motors. *Nature*, 417(6884), 25.
- Peper, F., Lee, J., Carmona, J., Cortadella, J., & Morita, K. (2013). Brownian circuits: fundamentals. *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, 9(1), 3.
- Qian, L., Winfree, E., & Bruck, J. (2011). Neural network computation with DNA strand displacement cascades. *Nature*, 475(7356), 368-372.
- Rapp, F. (2012). *Analytical Philosophy of Technology* (Vol. 63). Springer Science & Business Media.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioural and Brain Sciences*, 3(3), 417-424.
- Seeman, N. C. (2007). An overview of structural DNA nanotechnology. *Molecular Biotechnology*, 37(3), 246.
- Sha, R., Zhang, X., Liao, S., Constantinou, P. E., Ding, B., Wang, T., ... & Wu, G. (2005, October). Structural DNA nanotechnology: Molecular construction and computation. In *International Conference on Unconventional Computation* (pp. 20-31). Springer, Berlin, Heidelberg.
- Shanahan, M. (2015). *The Technological Singularity*. MIT press.

Shevlin, H. (forthcoming). Non-human consciousness and the specificity problem: a modest theoretical proposal. *Mind & Language*.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ... Graepel, T. et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419), 1140-1144.

Skillings, D. J. (2015). Mechanistic explanation of biological processes. *Philosophy of Science*, 82(5), 1139-1151.

Urban, M. W. (2012). Dynamic materials: the chemistry of self-healing. *Nature chemistry*, 4(2), 80.

Varghese, S., Elemans, J. A., Rowan, A. E., & Nolte, R. J. (2015). Molecular computing: paths to chemical Turing machines. *Chemical science*, 6(11), 6050-6058.

Von Neumann, J. (1956). Probabilistic logics and the synthesis of reliable organisms from unreliable components. *Automata studies*, 34, 43-98.

Wang, H. (1961). Proving theorems by pattern recognition—II. *Bell system technical journal*, 40(1), 1-41.

Wang, Z., Xiao, D., Li, W., & He, L. (2006). A DNA procedure for solving the shortest path problem. *Applied mathematics and computation*, 183(1), 79-84.

Wideman, J. G., Novick, A., Muñoz-Gómez, S. A., & Doolittle, W. F. (2019). Neutral evolution of cellular phenotypes. *Current Opinion in Genetics & Development*, 58, 87-94.

Woods, D., Doty, D., Myhrvold, C., Hui, J., Zhou, F., Yin, P., & Winfree, E. (2019). Diverse and robust molecular algorithms using reprogrammable DNA self-assembly. *Nature*, 567(7748), 366.

Yamamura, M., Hiroto, Y., & Matoba, T. (2001, June). Another realization of aqueous computing with peptide nucleic acid. In *International Workshop on DNA-Based Computers* (pp. 213-222). Springer, Berlin, Heidelberg.

Yogendra, K., Liyanagedera, C., Fan, D., Shim, Y., & Roy, K. (2017). Coupled spin-torque nano-oscillator-based computation: A simulation study. *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, 13(4), 1-24.

Yoshida, M., Hinkley, T., Tsuda, S., Abul-Haija, Y. M., McBurney, R. T., Kulikov, V., ... & Cronin, L. (2018). Using evolutionary algorithms and machine learning to explore sequence space for the discovery of antimicrobial peptides. *Chem*, 4(3), 533-543.