Normativity and the Intelligent Matter Framework

Barnaby Crook

Abstract An emerging field at the intersection of artificial intelligence, materials science, physical chemistry, and nanotechnology targets the creation of *intelligent matter*. However, despite demonstration of interesting behaviors in molecular systems and soft materials, the production of truly intelligent matter remains elusive. To address this, Kaspar and colleagues articulate a theoretical framework purporting to guide and evaluate the progress of the field. In this chapter, I argue that the absence of a *normative criterion* by which to assess the behavior of proto-intelligent material systems prevents the framework from adequately serving its guiding and evaluative functions. Without a clear specification of what kinds of behaviors count as intelligent, researchers are unlikely to develop material systems that exhibit them. I propose to augment the IMF with a normative criterion requiring purportedly intelligent systems to implement concrete behaviors which contribute towards higher-order goals in virtue of the coordinated activity of their functional elements.

1. Introduction

The *intelligent matter* research program aims to develop physical materials which implement intelligent behavior [1], [2], [3]. To coordinate this effort, Kaspar and colleagues formulate the *Intelligent Matter Framework* (IMF), which classifies material systems on the basis of which functional elements they possess.[1] Only by possessing the right set of functional elements can a system ascend to the category of *intelligent.* However, as I will show, this framework lacks *normativity.* That is, the IMF does not proffer any *normative criterion*[2] by which to judge patterns of behavior more or less intelligent. Absent such a criterion, I contend, any framework for intelligence will be inadequate. For one thing, theoretical terms such as *adaptivity* become divorced from their ordinary meanings. One cannot distinguish a material system's being *adaptive* from merely variable without some normative criterion by which to judge whether patterns of stimulus-induced dispositional change are beneficial. More fundamentally, without a normative standard dictating what counts as intelligent at the level of behavior, the IMF risks losing touch with the target phenomena that motivated

[1] I ascribe the name *Intelligent Matter Framework* to Kaspar and colleagues' work; it is not used in their original article.
[2] By *normative criterion* I mean any evaluative standard by which a system's behavior may be judged acceptable or unacceptable (in a dichotomous sense) or better or worse (in a continuous sense). No connotations to richer human domains of values and ethics are implied [4].

the research program in the first place. As such, I will argue, considerations of normativity are essential for any theoretical framework for intelligence [5], [6], [7], [8], [9], [10].

I critically analyze Kaspar and colleagues' theoretical framework for the intelligent matter research program from a philosophical perspective. I suggest that the IMF serves two complementary functions: i) an *evaluative* function, assessing progress towards the goals of the research program, and ii) a *guiding* function, which sheds light on how such progress might be achieved. I first highlight the lack of a normative criterion in Kaspar and colleagues' presentation of the IMF, before elaborating on the claim that this omission is problematic. In particular, I argue that without a normative criterion with which to judge behavior, the framework fails to adequately serve its functions. On the one hand, its evaluations fail to track intuitive judgements regarding behavioral sophistication. On the other hand, material systems which follow its recommendations may not exhibit any intelligent behavior at all. To explore how this might be rectified, I elaborate on the concept of normativity, distinguishing between the *source* of a normative criterion and how normativity is *constituted* in a physical system. I then discuss how normativity should be integrated into the IMF. I suggest that, in light of the intelligent matter research program's commitments to interdisciplinarity, practical applicability, and decentralized information processing, the IMF should be augmented with a normative criterion requiring material systems to implement extrinsically specified concrete behaviors which serve higher-order goals in virtue of the coordinated activity of their functional elements.

The structure of the chapter is as follows. In section 2, I present the IMF, describing how it divides material systems into classes based on their functional elements. In section 3, I focus on the absence of normativity in the IMF, arguing that the lack of a normative criterion leaves the IMF poorly suited to fulfilling its evaluative and guiding functions. In section 4, I cover normativity in more detail, clarifying the distinction between the source of a normative criterion and how it is constituted in a system. In section 5, I discuss which notion of normativity is best suited for integration into the IMF. I then conclude the chapter.

2. The Intelligent Matter Framework

In this section I introduce the IMF as described by Kaspar and colleagues in their 2021 article [1]. Before doing so, however, I briefly motivate focusing on this particular articulation of the intelligent matter research program. Since the aim of this chapter is to address conceptual issues arising in the definition and evaluation of intelligence, it is necessary to engage with a presentation explicitly delivered in conceptual terms, like the IMF. Further, research which falls under the purview of *intelligent matter* is conducted across a broad range of technical scientific disciplines, each with idiosyncratic terminology and research foci. What distinguishes the intelligent matter research program as a whole from any one discipline is its *integrative* nature. It is the challenge of combining diverse strands of technical work into a coherent research agenda that Kaspar and colleagues' IMF purports to address. And that

challenge has, I will argue, particular consequences for how normativity relates to the IMF. Finally, focusing on the IMF usefully restricts the scope of the chapter.

2.1 Introducing the Intelligent Matter Framework

The IMF classifies material systems on the basis of their internal functional complexity. In particular, the classification is sensitive to the presence of four specific functional elements: sensors, actuators, memory, and an interaction network. These elements are defined as follows. Sensors detect external stimuli, such as heat or force, by implementing an energy transformation (e.g., converting heat into electrical potential). Actuators produce a response, such as a physical or chemical change, when external stimuli are detected. Memory is any element in a system with the ability to retain information over a temporal window such that it can be exploited to modulate behavior later. Finally, an interaction network is constituted by signal pathways through which sensor, actuator, and memory elements can interact with one another, enabling feedback and other action-relevant information to be utilized. With all four of these functional elements in place, argue Kaspar and colleagues, a material system will be capable of sophisticated information processing and learning from its environment [1, p. 347].

Kaspar and colleagues use the functional elements described above to classify material systems. The most basic kind of system is *structural matter*. It does not possess sensors, actuators, memory, or an interaction network. Therefore, once a structural material system has been synthesized, it cannot change its properties. This means that, although such a system may have numerous functions and significant structural complexity, it is inherently unable to implement any behavior that requires adaptation, learning, or behavioral flexibility. The second class of material system Kaspar and colleagues describe is *responsive matter*. Systems belong to this category if they have sensors and actuators but no memory or interaction network. Because responsive material systems lack memory, the response their actuators produce when particular stimuli are detected by their sensors is always the same [1, p. 346]. The third class of material system is *adaptive matter*. The IMF classifies systems as adaptive if they augment the sensing and actuating capacities of a responsive system with an interaction network. Such a network allows a material system to implement feedback by altering its internal states and, consequently, its behavioral responses (e.g., in response to repeated presentation of a stimulus). Since adaptive material systems exhibit state-dependent responses to stimuli, the principles governing their behavior can become exceedingly complex (see [3] for discussion). The final and most sophisticated class of material system is *intelligent matter*. In addition to the functional elements required for a system to count as adaptive, an intelligent system must possess long-term memory storage. This enables previously sensed or self-generated information to be retained over time and deployed when needed. As well as guiding future action by modulating the behavior of actuators, feedback information retained as memory can be used to alter a system's internal states, implementing a form of learning.

2.2 *Preliminary Comments on the Intelligent Matter Framework*

I start by making three preliminary observations about the IMF. First, as a terminological matter, Kaspar and colleagues reserve the term *intelligent* for the final class of systems, stating that "if one of the key functional elements is lacking, then, according to our definition, the material is not considered intelligent" [1, p. 349]. Despite this assertion, the IMF can be interpreted as specifying discrete gradations of intelligence, with the four classes serving as distinct levels. Ultimately, the difference between these two perspectives is more semantic than substantial. For clarity, I will stick with the authors' usage moving forwards.

Second, the classificatory scheme of the IMF does not partition the space of possible material systems exhaustively. For example, where should a system possessing sensors and a memory, but no interaction network, be classified? Kaspar and colleagues discuss a soft matter system with such a functional repertoire and suggest that "such matter would not classify as adaptive matter owing to the lack of a network, [but nevertheless] goes beyond responsive capability" [1, p. 349]. While the existence of material systems which straddle the IMF's boundaries raises questions about its aptness, I put this issue aside for now.

Third, Kaspar and colleagues state that their notion of intelligence is not applicable to human beings or biological organisms more generally. For example, they write that "such matter which we term here intelligent does not show the same level of intelligence as would be understood in a psychological sense" [1, p. 345]. And, with respect to basal biological intelligence, they stress that the "key functional elements […] are not sufficient to enable the emergence of will or cognition, which distinguishes synthetic matter from intelligent living beings" [1, p. 347]. Such caution is admirable. Given the domain of research the IMF intends to support, eschewing psychological connotations seems appropriate. Indeed, such minimal strategies are common to researchers studying the evolution of intelligence [11], bacterial intelligence [12], plant intelligence [13], and artificial intelligence [14]. However, these disavowals raise the question of what relation the IMF's notion of intelligence *does* bear to those of other disciplines. One natural answer, not adopted by the IMF, is that diverse intelligence concepts are united by capturing systematic variation in *behavior* [6], [15]. That the IMF does not account for this will be the focus of my critique in section 3.3.

2.3 *Guide and Evaluate: The Two Functions of the Intelligent Matter Framework*

To critically discuss the IMF, it is necessary to consider what it hopes to achieve.[3] According to Kaspar and colleagues, they set out to provide "a roadmap towards intelligent matter" [1, p. 353] and to coordinate a "concerted, interdisciplinary and long-term research effort" [1, p. 353]. In other words, by specifying requirements for different levels of material intelligence, the IMF is intended to constitute a coherent and unified framework which can evaluate and

[3] This is a pragmatic approach which eschews metaphysical worries, e.g., about what intelligence *fundamentally* is or whether simple material systems can *really* possess it. Such questions, while plausibly interesting, are not within the stated remit of the IMF.

guide a distributed and diverse set of research practices. I suggest that this goal can be factored into two closely related functions which I describe below.

On the one hand, the IMF has a *guiding* function, serving to specify how the intelligent matter research program should aim towards implementing intelligence in material systems. According to Kaspar and colleagues', no extant system satisfy their criteria for intelligence [1, p. 353]. As such, the IMF's more demanding classifications serve as a lodestar for researchers to navigate towards. This role is explicit in the authors' concluding call to action: "…we must proceed from learning matter to truly intelligent matter, which receives input from the environment via sensory interfaces, shows a desired response encoded via embedded memory and artificial networks, and can respond to external stimuli via embedded transducers" [1, p. 353]. On the other hand, the IMF also has an *evaluative* function, enabling progress towards the goals of the intelligent matter research program to be assessed. To achieve this, specific material systems are examined and sorted according to the IMF's classificatory scheme. For example, Monreal Santiago and colleagues [16] developed a synthetic molecular system exploiting positive feedback to enable self-replication. According to the IMF, the system's possession of sensing, actuating, and interaction elements render it *adaptive* matter, but its lack of a long-term memory element precludes it being classed as *intelligent*.

Clearly, the guiding and evaluative functions of the IMF are intertwined. By evaluating extant systems and specifying which functional elements they are lacking, the future work required to develop intelligent matter is made apparent. To sum up, it is with respect to the IMF's aptness to guide and evaluate that it ought to be judged.

3. The Missing Role of Normativity

In this section I argue that the IMF has a weakness. In particular, it fails to provide normative constraints on which kinds of *behaviors* count as adaptive and intelligent according to its criteria. I begin by explaining the sense in which the IMF lacks normativity. I then contrast the IMF with definitions of intelligence which *do* provide this kind of normativity, to provide further clarity on what exactly is missing. Finally, I argue that normativity is critical for both the evaluative and guiding functions of the IMF.

3.1 Normative Criteria as Behavioral Standards

Clearly, the IMF provides some criteria for assessing the intelligence of material systems, namely those described in section 2.1. To explain the sense in which normative criteria are missing, then, a terminological distinction is required. I will refer to the criteria delineated by Kaspar and colleagues in the IMF as *functional criteria*, since they refer to a system's possession of, or constitution by, certain functional elements (i.e., sensors, actuators, an interaction network, long-term memory). There is a straightforward sense in which these functional criteria can be considered normative, since they specify properties that systems

need to have (i.e., constitutive norms to which systems need adhere) in order to achieve particular classifications. However, what I mean to refer to by *normative criteria* are standards applied to the *behavior* of a system (i.e., what it does). For example, a thermostat is judged (prima facie) by its capacity to minimize the difference between the temperature of a physical system and a specified set-point. In this sense, normative standards function as *performance measures* by which a system's ability to implement target behaviors may be assessed [14]. Possessing the right combination of functional elements may be a pre-requisite for achieving some behavioral outcomes, but the achievement of most outcomes will typically require something over and above possession of those elements (e.g., the elements being organized such that their operation in typical environments reliably produces the relevant outcomes). To capture this distinction, when I speak of *normative criteria* moving forwards, I am always referring to standards by which systems' behavior may be evaluated, regardless of differences in the functional elements involved.

So then, to make my claim that the IMF lacks a normative criterion precise, it can be re-phrased as follows. According to the IMF, conditional on the set of functional elements that a given material system possesses, variation in the behavioral propensities of that system is immaterial to the class into which it should be sorted. That is, once we know which functional elements a system is endowed with, we do not need to know anything more about how those functional elements interact or which patterns of behavior they produce in order to know how intelligent that system is.[4] I now contrast the IMF with other frameworks of intelligence in order to clarify the typical role of normative criteria.

3.2 Normativity and Definitions of Intelligence

In their review of definitions of intelligence, Legg and Hutter synthesized over 70 proposals from psychology, cognitive science, and artificial intelligence into a single sentence: "Intelligence measures an agent's ability to achieve goals in a wide range of environments" [17, p. 9]. This broad and general definition brings two distinctions with the IMF to light. First, the synthesized definition explicitly invokes normative criteria through reference to "goals". While the notion of a goal remains imprecise in the definition, however it is spelled out, it will constrain what a system needs to *do* in order to achieve a certain degree of intelligence (I will have more to say on goals in section 4.2). This is exactly what the IMF lacks. Second, in contrast to the IMF, the synthesized definition specifies no *functional* criteria at all. That is, it does not place any requirements on *how* a given system achieves its goal. I do not suggest that the IMF is wrong to specify functional criteria. On the contrary, specifying such criteria enables the IMF to fulfil its guiding function (with the proviso that it will only function well as a guide under the condition that the criteria are well-motivated).

[4] As I will discuss in section 3.4, textual evidence casts suggest that Kaspar and colleagues did not intend this to be a property of their framework. Rather, the omission of an explicit normative criterion by which the behaviour of purportedly intelligent systems may be judged is more likely an oversight.

Instead, my claim is that, especially for the IMF to fulfil its evaluative function, it needs normative criteria *in addition to* functional ones.

Appeals to normative notions in the definition and evaluation of intelligence are ubiquitous. Normativity in Legg and Hutter's definition of intelligence is conferred by the use of the word "goal". However, there are numerous ways for normativity to enter a definition or framework for intelligence.[5] For example, Russell and Norvig structure their approach to AI around the idea that "some agents behave better than others", which "leads naturally to the idea of a rational agent—one that behaves as well as possible" [14, p. 95]. Through invoking notions of *better*, *worse*, and *best* behaviors, their concept of rationality becomes normative. One might argue here that the approach relevant to traditional AI agents should differ from that for material systems. However, Russell and Norvig define an *agent* as "anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators" [14, p. 96]. As such, their approach is clearly minimal enough to be compatible with intelligent matter (indeed, even responsive matter is covered). A non-exhaustive list of terms, common to definitions of intelligence, which (usually) connote a normative standard include: rationality [18], problem-solving [19], indistinguishability from human beings [20], adaptability/adaptivity [10], task-performance [7], and more besides [6], see [15], [21]. All definitions which invoke such notions incorporate normativity.

One term mentioned above, *adaptivity*, appears in the IMF and is important to discuss in greater detail. One might think the worry about normativity is forestalled by the IMF's inclusion of this term. However, there is a critical contrast between how the IMF uses the word *adaptive* and how it is typically applied in biological and AI settings. Recall that in the IMF a material system is adaptive if it possesses sensors, actuators, and an interaction network [1, p. 346]. In biology, in contrast, adaptivity is normally defined in terms of how a trait contributes to the inclusive genetic fitness of the organism that possesses it [22]. For example, it is adaptive for cacti to grow spines because they deter herbivores who might otherwise eat the cacti, preventing them from replicating and propagating their genetic material [23, p. 26]. We can readily imagine a scenario in which a cactus' growing of spines depends upon the presence of some environmental trigger, such as a particular quantity of rainfall. In such a case, the cactus is making use of sensors, to detect the presence of rain, actuators, to initiate the growth of spines once the threshold is reached, and a network, through which the signals from sensation to actuation are propagated. However, from a biological perspective, such behavior only counts as adaptive if the inclusive genetic fitness of the cactus, conditional on its spine-growing behavior, is greater than its inclusive genetic fitness in a counterfactual world in which the spine-growing behavior was absent. In other words, biological adaptivity tracks *beneficial contribution*, not functional complexity. While

[5] I do not distinguish sharply between *definitions* and *frameworks* here. Roughly speaking, the IMF is a framework composed of a set of hierarchically arranged definitions. However, the evaluative and guiding functions of the IMF may also be played by singular definitions [10].

particular forms of complexity might be necessary to attain certain benefits [11], unless they actually *do* attain those benefits, at least in expectation, they are not considered adaptive.[6]

3.3 Normativity is Crucial for the Evaluative and Guiding Functions of the IMF

So far in this section, I have specified what is meant by the claim that a normative criterion is absent from the IMF as stated and shown that this feature marks the IMF out as unusual among attempts to define or conceptualize intelligence. I now argue directly that augmenting the IMF with an explicit normative criterion is required for it to adequately serve its evaluative and guiding functions.

Pei Wang contends that a working definition for artificial intelligence (like the IMF) ought to adhere to four requirements (initially proposed by Carnap in his discussion of probability [24], [10] (see [6] for a related approach). Those requirements are as follows: i) *similarity* to the way the term is commonly used, ii) *exactness*, iii) *fruitfulness*, and iv) *simplicity*. The similarity requirement prevents an academic definition of a term from becoming divorced from its ordinary meaning. For example, we could stipulatively define intelligence for material systems as a measure of, say, conformational flexibility. But this would leave us with a term utterly failing to capture the sense of intelligence that motivated the need to define it in the first place. The exactness requirement demands that a definition be precise enough to aid scientific discourse. In the context of intelligent matter, exactness is critical for the IMF to serve its evaluative function.[7] The fruitfulness requirement demands that a definition be informative enough to serve a practical role. For the IMF, this equates to fulfilling its guiding function. As Wang states, researcher definitions of intelligence typically do not describe "something that already fully exists, but something to be built" [10, p. 5], a description that applies clearly to the IMF. The simplicity requirement demands that a definition be as simple as possible, given that it satisfies the first three requirements.[8]

I suggest that an explicit normative criterion is crucial for any definition or framework of intelligence to adhere to the similarity requirement. As Kryven and colleagues put it, "attributing intelligent behavior relies on interpreting an agent's actions as … reasonable in the first place" [26, p. 15]. In other words, the purpose of ascribing intelligence to entities is as a means to explain and predict systematic differences in their observed behaviors. In the human domain, some people are able to complete cognitive tasks that others cannot. As Sternberg notes [27, p. 308], it is with respect to such observations that the ascription of

[6] The AI use of *adaptivity* is essentially parasitic on the biological notion e.g., [10, p. 19] but with inclusive genetic fitness exchanged for a quantitative performance measure.
[7] As illustration of the perils of inexactness, consider Legg and Hutter's synthesized definition of intelligence (section 3.1). By invoking undefined notions of *agents*, *goals*, and *environments*, this definition leaves much up to interpretation should one attempt to apply it to a real-world material system (see [25] for an attempt to make that definition exact).
[8] There is a tension between this requirement and the third, since simple definitions may lack the detail required for informativity, but discussing this trade-off is beyond the scope of this chapter.

differing degrees of intelligence to individuals has explanatory value.[9] For this reason, any framework of intelligence that fails to track observed variation in relevant behavioral outcomes is insufficiently similar to the ordinary concept [6].[10]

While it may not immediately seem catastrophic for the IMF to fail to adhere to one out of four requirements, the value of providing a sufficiently *exact* and *fruitful* definition can only be realized given *similarity* is satisfied. Otherwise, the way in which systems are evaluated, even if precise, will ultimately fail to track our intuitive judgements. Furthermore, systems which follow the guidance of the IMF may fail to demonstrate behavior which would independently strike us as intelligent. If my critique is right, there should be possible systems which would satisfy the IMF's criteria for adaptivity and intelligence but fail to produce intelligent behavior. To that end, consider a system equipped with a sensor that periodically detects (external) temperature, a memory system that stores experienced temperatures as sequences (of length 3, say), an actuator that can trigger the rotation of the system (like a bacterial flagellar motor), and a network mechanism connecting these functional elements. The network mechanism evaluates the most recent trajectory of temperatures and, if it matches a stored sequence, signals the actuator to trigger the rotation of the system. The implementational details of this imagined case are not important.[11] The point is that this system would, in virtue of its possession of all functional elements specified by the IMF, constitute intelligent matter. Despite that, it is hard to see how such a system could possibly implement behavior that would count as intelligent by a practical normative standard.

As well as failing to track intuitive judgements in cases where functional complexity is put to unintelligent use, the IMF also overlooks that what a system *does* is limited in important ways by numerous factors beyond which functional elements it possesses. These factors include i) the variety of things it can sense, ii) the range of actions available to it, and iii) the fit between the behavioral policy it implements and the conditions of the environment [30]. It is true that an enlarged functional repertoire expands the space of possible behaviors. However, merely possessing such a repertoire does not ensure that the behaviors plucked from that vast space belong to the subset of valuable ones. For this reason, a measure of intelligence that focuses entirely on functional criteria is bound to overrate systems which do useless things for sophisticated reasons. In order to cleave to the ordinary sense of the term *intelligent*, then, the IMF must augment its specification of the functional elements necessary for intelligence with some kind of normative criterion for the behaviors material systems implement.

---

[9] Indeed, one might view intelligence as no more than a useful shorthand for speaking about these observed differences in capabilities and behaviours (see [28] for such a view). Of course, there are more theoretically laden ways to interpret *intelligence*, with the IMF being an example. The point is that, to satisfy *similarity*, an explication of the concept ought to summarize observations in a compact and predictively valuable manner.

[10] This is not to say that the exact outcomes over which variation is tracked need to be the same as in the human case. As acknowledged in section 2.2, Kaspar and colleagues do not intend their notion of intelligence to be applicable to psychology.

[11] One can imagine such a system as a variant of Braitenberg's *Vehicle 5* [29].

Before I discuss normativity in more detail, I briefly examine textual evidence suggesting that Kaspar and colleagues implicitly invoke normativity when applying the IMF.

3.4 Implicit Normativity in Applying the IMF

Although normativity is absent from the explicit presentation of the IMF, it still factors into the judgements Kaspar and colleagues make about the intelligence of material systems. There are two ways this plays out. First, although the IMF only directly classifies systems with respect to *how* they implement behaviors (i.e., functional criteria), there are implicit criteria governing which range of behaviors a system must display to be deemed interesting enough to warrant examination (i.e., implicit normative criteria). Second, comments from Kaspar and colleagues imply that further details relating to how systems *implement* normativity are relevant to how intelligent they consider them to be. I address these points in turn.

Normative criteria implicitly govern which systems are candidates for classification as intelligent by the IMF. Consider Yu and colleagues' demonstration of self-swarming behavior in paramagnetic nanoparticles [31]. Due to their magnetic interactions, upon exposure to oscillating magnetic fields these particles display a range of pattern-forming collective behaviors. In their assessment, Kaspar and colleague state that, because the nanoparticles "rely on the input of an external programmer who manipulates the magnetic field… the particles do not show intelligent behavior by themselves" [1, p. 347]. Notice that this statement pertains to the *behavior* of the nanoparticle system as a function of environmental variation (i.e., input). It is because this pattern-forming behavior is only observed under environmental conditions which require close external control that the system is deemed less autonomous and, thus, less intelligent. Such a claim implicitly invokes a normative criterion by focusing on behaviors deemed interesting from the outside (e.g., collective motility). Without such a criterion, there would be no reason for the pattern-forming behavior, generated through controlled deformations of the particle swarm, to be distinguished from any other behavior. Similarly, He and colleagues [32] (described in [1, p. 350]) develop a system which, through a feedback loop involving the mechanical action of temperature-responsive hydrogel and the chemical action of exothermic catalytic reactions, maintains its temperature in a narrow range. In this case, thermoregulation is the behavior judged worthy by an implicit normative criterion. The failure to make these implicit judgements explicit may seem trivial. However, the hypothetical system considered in the previous section shows how this can manifest in unintuitive judgements. Further, since the IMF is supposed to guide research, leaving such criteria implicit is unsatisfactory (see section 5.4 for further discussion).

Second, Kaspar and colleagues suggest a further constraint, not prescribed by the IMF, on how norms should be *implemented*. In particular, they imply that the criterion by which a material system's behavior should be judged must, in some sense, be generated from *within* the system. For example, when assessing a swarm of autonomous robots that can assemble

into two dimensional patterns, Kaspar and colleagues state that "since an external programmer predefines the targeted shape and gives instructions in [the] form of an algorithm, the whole group of robots is not intelligent according to our definition" [1, p. 346]. It is not clear how the authors derive this judgement from their definition of intelligent matter, which does not mention goals or the achievement of target states, whether defined externally or not. However, this comment (and the similar comment about the swarming paramagnetic nanoparticles) suggests that Kaspar and colleagues intend the IMF to limit the ascription of intelligence to systems which, in some respect, generate their own normative criteria. Call this an *intrinsicality* condition on normativity. I discuss this requirement further in section 5.2.

4. Aspects of Normativity

In section 3 I observed that the IMF lacks a normative criterion and argued that this omission is problematic. To rectify this situation, I suggest, *some* notion of normativity ought to be explicitly incorporated into the IMF. This should serve to make the framework sensitive to behavioral capacities in a way that tracks the ordinary notion of intelligence, thus satisfying the similarity requirement (see section 3.3) and enabling the IMF to serve its evaluative and guiding functions effectively. However, for this to be possible, clarity around *which* notion of normativity ought to be incorporated is required. To that end, this section unpacks the concept of normativity.

I suggest that questions about normativity can be divided into two distinct categories.[12] One set of questions pertains to the *source* of the normativity. That is, where does the normative criterion originate? By what process does it arise? Another set of questions pertains to the *constitution* of the normativity. That is, how is the normativity *physically instantiated* in the system? Are there explicitly represented goal states against which actual states can be compared? Or is the normativity governed by the organization of the entire system? Let us take a closer look at each of the aspects of normativity before discussing their relation to the IMF.

4.1 Aspects of Normativity: Source

First, let us consider *sources* of normative criteria. That is, where do such criteria come from? The source of normativity may be considered from an ultimate or proximal perspective [33], [34], [35], [36]. In the case of biological systems, the ultimate source of a normative criterion is the process of evolution by natural selection [23], [37]. It is inherent to

[12] To structure this section, I take inspiration from Rahwan and colleagues' typology of questions for machine behaviour [33], itself derived from Tinbergen's foundational work in ethology [34]. Questions about what I call the *source* of normativity roughly map on to Tinbergen's *historical/developmental* questions, while questions about the *constitution* of normativity roughly map on to Tinbergen's *contemporary/slice-in-time* questions. However, I intend the relation to Tinbergen (and Rahwan et al.) to be one of rough inspiration, not isomorphism.

natural selection that success for an individual is constituted by surviving and reproducing, with the value of traits and behaviors derived from their contribution to these core objectives [38]. Without the differential replication of variant biological forms, captured formally by the metric of inclusive genetic fitness, there could be no criterion for the success of individuals and thus no normativity [39]. Proximal sources of normative criteria for biological organisms are the specific challenges posed by their environments ([40], see [41] for detailed discussion of how normativity relates to biological function).[13] For example, the threat of predation by foxes imposes a normative criterion on rabbits by rendering their survival conditional on their ability to evade capture. On this view, the evolution of behaviorally and cognitively sophisticated organisms is a product of an increasingly diverse multitude of proximal sources of normativity, all generated from the same ultimate source of natural selection (see [11] for discussion).

For engineered systems, the ultimate source of normativity is challenging to specify. There is no consensus view on how to determine what constitutes appropriate functioning or overall success for an artifact (see [43] for philosophical discussion). However, to a first approximation, we can follow the example of biology and characterize success in terms of spread and prevalence (as in [33]). Translating these ideas to the world of artifacts, we are faced with an ultimate criterion of widespread manufacture, deployment, and use.[14] Another way to frame this is that the ultimate source of normativity for artifacts, by analogy to biological evolution, is cultural evolution [44].[15] Though there is significant controversy about how best to describe variation and selection in a cultural evolutionary framework [45], the relevant dynamics likely involve a complex combination of social, technical, and biological factors [33], [46]. Moving onto the proximal sources of normativity, once again, we are faced with variety. In some cases, such as formal verification in software engineering [47], [48], normative criteria for the behavior of engineered systems may be formalized and explicated precisely. However, in other cases, the criteria by which system behavior is evaluated may be challenging to state exactly [49]. For example, consider the recent explosion in the prevalence and usage of large language models [50]. While the formal measure of perplexity captures how well models such as ChatGPT perform at their training objective [51], their commercial success is a function of how useful people find them, a metric which cannot be easily operationalized prior to releasing the model as a product. The use of performance on task benchmarks as proximal normative criteria has been a topic of critical discussion within the AI and machine learning community (e.g., [5], [21]). This

[13] It would be beyond the scope of this chapter to discuss the intricacies of the philosophical debate on biological functions. However, see [42] for an attractive pragmatic and pluralist view.

[14] This relates to the etiological view of artifact functions [43]. However, I do not take describing the source of normativity for engineered artifacts to require commitment to such a view (or any view at all) about artifact functions.

[15] Note that there is ongoing debate about the nature, prevalence, and explanatory power of cultural evolution (as well as whether it is an *evolutionary* process in a literal, analogical, or metaphorical sense). However, the basic observation that technical artifacts differ in ubiquity and persistence is undeniable.

reflects the difficulty of selecting proximal normative criteria which lead to desirable outcomes [52], [53], [54] (see also section 5.4).

Commented [bc1]: Cut or change?

4.2 Aspects of Normativity: Constitution

Next, let us consider how normativity may be *constituted* in a system. That is, what physical arrangement of matter *implements* normativity within a system? Here, two broad categories can be distinguished. First, some systems explicitly represent goal states or objectives such that they can evaluate occurrent states with respect to them [55]. Call this category *explicit representation*. Explicit representation of goals is commonly associated with the symbolic approach to AI spearheaded by Newell and Simon [56]. Newell and Simon developed so-called *expert systems* which operated via backchaining, i.e., given a goal state and an occurrent state, searching for a trajectory of intermediate states such that the goal state is reached [57]. More generally, approaches to AI that focus on search and planning typically involve explicit representation of goals [14]. Control systems which explicitly represent their control variables (i.e., those whose values they maintain within a specific range) also belong to this category [58], [59]. Additionally, though the details are controversial, it is widely believed in cognitive science that human beings explicitly represent goals in some manner [60]. Whether or not the same is true in various species of nonhuman animals is a matter of ongoing debate [61], [62], [63].[16]

Commented [bc2]: Only two? What about systems which fail to meet the criteria for either group (i.e., reflex agents)?

The second broad category concerning the constitution of normativity is *holistic arrangement*. Here, there is no (canonical) way to localize the normativity which governs the behavior of a system. Rather, it is the *entire* physical arrangement of the system which serves to dictate that its behavior conforms to a particular norm (these are what [55] call 'goal-seeking systems', or what [65] calls the 'systems view'). Again, both biological and engineered systems may fall into this category. Indeed, arguably, every biological system is normative in virtue of its holistic arrangement, since it is the way the system is arranged (in both space and time) which constitutes its identity [62], [66]. Broadly speaking, this is the position taken up by proponents of the autopoietic view of biological systems and their intellectual descendants [67], [68], [69].[17]

When it comes to engineered systems, they do not (yet) possess the dynamic cohesion or behavioral flexibility characteristic of biological organisms [70], precluding the kind of autonomy exhibited by living beings. However, there is still clearly a class of human-designed material systems for which their overall functional organisation imbues them with a weaker kind of purposiveness. Indeed, better understanding how this subset of engineered

[16] It is beyond the scope of this chapter to enter into detailed analysis regarding what is required for something to count as a representation of a goal. Indeed, it is plausible that no sharp dividing line separates biological systems which do and do not explicitly represent goals [64]. If this is so, it seems equally plausible that artificial systems, including those developed in materials science, may also blur the boundaries.

[17] Note that commitment to a holistic arrangement account of biological normativity is compatible with believing that (some) biological organisms explicitly represent target states. However, the two accounts are mutually exclusive as explanations for the achievement of any *particular* state.

machines, so-called *servomechanisms*, implemented goal-directed behavior was a central goal of the cybernetics movement of the 1940s and 50s [9], [71], [72]. To anticipate a possible misunderstanding, note that holistically arranged normativity is not trivially ascribable to every human-designed artifact, merely in virtue of the fact that their intentionally designed characteristics contributes to their utility, spread, and application. The critical difference between normative systems such as, say, a heat-seeking missile, and non-normative ones such as, say, a clock, is that the former but not the latter have the propensity to reach or maintain particular desirable states despite environmental perturbations ([23] views this as an aspect of goal-directedness, while [65] calls this property 'persistence').[18] Systems which do not exhibit any flexible behavior that reliably reaches or maintains particular target states do not implement normativity.

Commented [bc3]: Check this sentence. Did cyberneticists really focus on holistically organized machines?

With the source and constitution of normativity in mind, let us turn to considering which senses of the term could be usefully incorporated into the IMF.

5. Normativity and the Intelligent Matter Framework

To recap, section 2 presented the IMF, posited that it aims to serve evaluative and guiding functions, and noted that it omits explicit normative criteria. Section 3 discussed this omission in detail, arguing that it prevents the IMF from adequately fulfilling both its evaluative and its guiding function. Section 4 provided a brief analysis of normativity, suggesting that the *source* from which normativity derives and how normativity is *constituted* in a given system are distinct, important considerations. This section turns to how normativity should be inserted into the IMF in a way that accords with the theoretical commitments of the intelligent matter research program. In particular, we can ask: i) which source of normativity is most suitable for the purposes of the IMF? And: ii) what requirements for the constitution of normativity, if any, ought the IMF to impose?

5.1 Sources of Normativity for the IMF

Recall that the source of a normative criterion refers to where it comes from. That is, what is responsible for generating the evaluative standards by which a material system's behavior should be judged? As an initial question, we can consider whether it is the ultimate or proximal source of normativity with which the IMF should be concerned. This has a straightforward answer. Any attempt to dictate the *ultimate* source of normativity for engineered artifacts would be hubristic folly. Which engineered artifacts and material systems end up being reproduced, distributed, and deployed (i.e., the ultimate source of normativity) depends upon myriad factors that could not be effectively enumerated, let alone controlled, by any research effort. By elimination then, it is the *proximal* source of

[18] A precise explication of the degree to which a system needs to exhibit this form of persistence to count as normative is outside the scope of this chapter. Personally, I am inclined towards a gradualist perspective on which there is no definite boundary.

normativity that the IMF should be concerned with. So, the question becomes: how should the IMF impose local normative standards to better fulfil its functions?

Two general strategies can be delineated.[19] On the one hand, the IMF may augment its functional requirements with a specific normative criterion that is applicable across cases. For example, one option would be to follow Legg and Hutter's synthesized definition of intelligence discussed in section 3.2 ("the ability to achieve goals in a wide range of environments"). In order to give this definition teeth, Kaspar and colleagues would need to define "goals" and "environments" concretely enough to quantify the range of environments systems operate in. They would also need to consider how to integrate a measure of this quantity with the discrete functional requirements specified by the IMF in its current form. For instance, each class of matter delineated by the IMF could be associated with a threshold of the measured quantity such that a system can only count as, say, adaptive, if it achieves goals across sufficiently many environments. This approach has the potential to endow the IMF with the power to fulfil its evaluative and guiding functions, ruling out degenerate systems and providing a clear vision of intelligent behavior. However, enacting this strategy may be challenging given the breadth and diversity of the intelligent matter research program. Devising a criterion equally applicable to self-organizing nanoparticle assemblies, soft matter systems, and distributed neuromorphic computing systems may not be possible. For example, Legg and Hutter's definition may struggle to capture the fact that some intelligent matter research aims to replicate the flexibility of general computation (e.g., [73]) while others aim for the robust implementation of narrow behaviors (e.g., [31]).

The other avenue available to the IMF is a *devolved* approach which transfers responsibility for specifying normative criteria to the researchers developing purportedly intelligent material systems. This strategy comports with the spirit in which the article presenting the IMF is written. For example, consider He and colleagues' homeostatic mechanochemical system, in which "a continuous feedback loop between an exothermic catalytic reaction […] and the mechanical action of [a] temperature-responsive gel results in an autonomous, self-sustained system that maintains temperature within a narrow range" [32, p. 350]. Adopting the devolved approach would mean deferring to He and colleagues with respect to the normative standard by which the system is evaluated. In this case, the system would (presumably) satisfy the normative standard imposed by the researchers in virtue of its thermoregulatory behavior. This strategy has the advantage of adaptability. It can capture whatever normative criteria are deemed relevant by particular scientists pursuing the varied research programs which fall under the banner of *intelligent matter.* As a drawback, however, the devolved strategy provides less in terms of both guidance and evaluation than the specification of a particular measure (see section 5.4 for further discussion of this point).

[19] For clarity I present these strategies as discrete options. However, in reality there is a continuum on which strategies may be placed.

Since there are pros and cons to both the specific approach and the devolved approach, I suggest a compromise position that augments the IMF with the following normative criterion:

NC: In order for a material system to be classification as intelligent (or adaptive, responsive, or structural), it must be endowed with the requisite functional elements *and*, through the coordinated activity of those functional elements, robustly implement concrete behaviors (e.g., the maintenance or achievement of one or more target states) which contribute towards a higher-order goal.

This criterion reflects a partially devolved strategy. It leaves both the concrete behaviors a system implements and the higher-order goal towards which the system's behavior contributes open to interpretation. However, it does still impose general constraints which rule out degenerate systems as described in section 3.3. In particular, only systems which achieve specified target states of some kind, be they internal states of the system or states of the environment, can count as intelligent. That systems must do this *robustly* ensures that some degree of persistence in the face of perturbation is required. Further, that these states must contribute to some higher-order goal brings the IMF's use of the term *adaptivity* into alignment with other disciplinary approaches to intelligence. [20] Finally, the requirement for the behavior of the system to be implemented by the coordinated activity of its relevant functional elements ensures that the functional and normative criteria are integrated into a coherent whole.

Practically, application of the augmented IMF would involve specifying which concrete behaviors constitute the norm for a particular material system (e.g., maintaining system temperature in a particular range) and what higher-order goal those behaviors contribute to (e.g., the persistence of the system).

5.2 Constitution of Normativity for the IMF

Let us now turn to what prescriptions, if any, the IMF might make regarding how normativity should be *constituted* in purportedly intelligent material systems. In particular, is one of the two categories delineated above, explicit representation and holistic arrangement, more fitting than the other? We should first note that the IMF, in virtue of supplying a functional parts list necessary for systems to be designated as responsive, adaptive, or intelligent, already constrains how *any* behavior ought to be implemented. Thus, we can begin by considering whether these constraints suffice to commit the IMF one way or another regarding the implementation of normativity. That is, may systems possessing the relevant functional elements adhere to a behavioral norm in virtue of *either* explicitly represented goal states *or* holistic arrangement, or does that functional profile already exclude one of those possibilities? It is clear that both possibilities remain open to the IMF. Goal states may

[20] Indeed, the system I imagined in section 3.3 could be rescued from the charge of degeneracy by specifying the higher-order goals to which its behavior contributes.

be explicitly represented in a long term memory or interaction network, but flexible goal-seeking behavior is possible without such representation [55], [65]. As such, the IMF is neutral with respect to how exactly those functional elements might be constrained to interact in ways which reliably lead to desirable states.

However, although both explicit representation and holistic arrangement are logically consistent ways for the IMF to bolster its functional requirements, there are further considerations, touched upon by Kaspar and colleagues, which favor holistic organisation. I will consider two such points. First, as noted in section 3.4, Kaspar and colleagues invoke an *intrinsicality* condition in their evaluation of the swarm behavior of autonomous robots and paramagnetic nanoparticles. In each case, the systems in question are deemed adaptive but not intelligent in virtue of their requiring an external programmer to define the target shapes of the swarm [1, pp. 346–347]. This implies that if the target shapes had *not required* this external programming, the systems would have been deemed intelligent. This interpretation is supported by Kaspar and colleagues' suggestion that natural self-organizing swarms (which are not externally programmed), such as insect colonies and schools of fish, are "often considered to exhibit features of intelligent behavior" in virtue of being "coordinated in a decentralized manner" (p.346). This brings us to *decentralization*, which is the second consideration. Developing decentralized intelligent systems is an important ambition of the intelligent matter research program. As Kaspar et al. state, the ideal combination of their favored functional elements would "form functional processing continua, which do not require a centralized processing unit, but rather provide the capability for local and distributed information processing" [1, p. 345]. If we take the authors at their word, the normativity of intelligent material systems should be constituted in a decentralized but intrinsic manner.

What does it mean for an engineered system to exhibit decentralized, intrinsically normative behavior? If the behavior that the system displays is decentralized, then it is not being governed by a central information processing unit. This implies that the system cannot be explicitly representing its goal state in a dedicated location, since that location would constitute a *de facto* locus of control.[21] If this reasoning is sound, then a decentralized system must implement normativity, if at all, through its *holistic arrangement*. Next, let us contend with the notion that this normativity should be, in some sense, *intrinsic*, that is, not dictated by an external source like a programmer. How should we interpret this requirement? One possibility would be that the system must exhibit the strong sense of holistic arrangement, as explicated by researchers inspired by the autopoietic tradition. On this view, a biological system is special because "the conditions of existence of its constitutive processes and organization are the norms of its own activity" [69, p. 34]. This means that an "interaction or process is detected as "bad" or "good" by and for the very

[21] In theory, nothing prevents a decentralized system from representing a goal state in an inert way, disconnected from the functional organisation which brings about that goal state. However, such a system would not satisfy typical criteria for explicit representation e.g., [55, pp. 128–9]

system (not by and for an external observer)" [69, p. 35]. Ruiz-Mirazo and Moreno capture a coherent sense in which a system can have intrinsic norms. However, importing this into the IMF would be at odds with the stated focus on commercial engineering applications like intelligent clothing and neuromorphic computing, as well as the explicit disavowal that developing intelligent matter implies producing willful living systems (see section 2.2). Further, if the IMF adopted this demanding intrinsicality requirement, then its research goals would collapse into those of synthetic biology and artificial life [74], [75].

This leaves the IMF in a bind. Adopting a standard of strong intrinsicality seems contrary to the spirit of the intelligent matter project. On the other hand, there may be no other meaningful sense in which a system which does not select among explicitly represented goal states can exhibit intrinsic normativity. As Ann Sophie Meincke puts it, "it is not clear in what sense the behaviors of systems [like] thermostats or torpedoes could be guided by intrinsic goals, given that there is no sense in which such systems can be considered as themselves setting these goals" [76, p. 286]. If this argument is right, then conditional on a system's normativity being constituted by its holistic arrangement, either it pursues the biological imperative to maintain its own dynamic organisation (in the strong sense demanded by the autopoietic tradition), or it cannot be intrinsically normative. Thus, it may not be possible for the IMF to simultaneously retain all of its stated desiderata. One of the following – holistic arrangement, intrinsic normativity, practical applicability – must be jettisoned. In my view, the intrinsic normativity requirement is the most sensible candidate for rejection. While there is appeal to the notion that a material system which governs its own norms is more intelligent than one that does not, I cannot see conceptual space for a research endeavor which adopts the requirement of intrinsic normativity while retaining an identity distinct from that of already extant research programs in synthetic biology and artificial life. The project of developing a material system which is intrinsically normative in virtue of its holistic arrangement simply *is* the project of developing an artificial living system.

5.3 Partially Devolved, Extrinsic, Holistically Arranged Normativity for the IMF

We have examined numerous possible ways the IMF might be augmented with normativity. As for its source, I have argued that, given the heterogeneity of research that falls under the purview of *intelligent matter*, a partially devolved solution is best. This means that the extent to which a material system exhibits concrete target behaviors defined by the researchers investigating that system should be considered, explicitly, on a case-by-case basis. While this approach is deflationary, it is also pragmatic, avoiding the difficulty of delineating a specific, quantitative criterion for intelligent behavior that applies across diverse research contexts. When it comes to how normativity ought to be constituted, I have suggested that the set of preferences expressed in Kaspar and colleagues' article are mutually inconsistent. In particular, no research project can simultaneously develop material systems which are suitable for a diverse range of practical purposes (as opposed to synthetic living systems), intrinsically normative, and normative in virtue of their holistic arrangement. As such, for the IMF to capture and guide a coherent research agenda, I have argued that it

ought to drop the requirement for intrinsic normativity, allowing systems to be considered intelligent even if their target states are specified externally. This preserves an independent identity for intelligent matter, preventing a collapse into synonymy with synthetic biology and artificial life. Further, it enables the requirement for systems to implement intelligent behavior in a decentralized manner to be retained, further distinguishing the intelligent matter research program from AI approaches based on traditional computational paradigms [14].

5.4 Discussion

I have suggested that the normativity most apt for introduction into the IMF is partially devolved, extrinsic, and holistically arranged. However, it is worth considering this position's major weakness. There is a potential tension between the desideratum for practical applicability and the goal of producing intelligent matter. The devolved approach to normativity assumes that material systems which operate via the coordinated activity of the IMF's functional elements will be developed naturally in the process of targeting practical applications. However, historical precedent suggests such conjectures are on shaky ground. The history of AI is littered with shattered expectations about the degree of functional complexity required to implement particular tasks [77]. For example, in 1979 Douglas Hofstadter claimed that expert-level chess playing would require a generally intelligent system [78, p. 678]. Given how often such predictions have failed, the expectation that matter satisfying the IMF's criteria for intelligence will arise organically in pursuit of practically applicable systems should not be taken for granted. As such, the devolved approach could undermine the IMF's ability to serve its guiding function.

The point above highlights an important challenge for the IMF: identifying target behaviors which genuinely necessitate the incorporation of the functional elements specified by the IMF (see [6]). Kaspar and colleagues' article understandably focuses on the difficulties associated with implementing functional elements in material systems. However, as argued in section 3.3, possession of such elements does not, by itself, suffice for intelligent behavior. As such, I suggest, it would be valuable for proponents of the IMF to propose concrete behaviors or properties which can only be exhibited by systems satisfying the IMF's functional criteria.[22] Such behaviors would not constitute *necessary* conditions for intelligence and would not be part of the IMF itself. Instead, they would accompany the IMF as exemplars, providing additional guidance to the research field.

Before concluding, I will clarify the contribution of the foregoing discussion. Although I have offered a positive proposal, my suggestion relies on an interpretation of the goals of the intelligent matter researcher program, which may be flawed. Thus, I stress that the philosophical contributions here include i) the argument that normativity is relevant to the IMF, ii) a clarification of different things normativity can mean, and iii) analysis of the

[22] This is intended in an informal sense, i.e., given typical constraints on spatial, temporal, and computational resources.

relationship between various notions of normativity and the function of the IMF. Crucially, if my argument is sound and the analysis correct, the dependencies between the ambitions of the intelligent matter research program and the IMF's approach to normativity will remain, even if researchers prefer adopting an alternative strategy in light of those dependencies. For example, the interdisciplinary breadth of the research program may be a less significant barrier to devising a specific, quantitative normative criterion than I suggested. If this is so, then developing such a criterion would surely be worthwhile. Similarly, researchers may choose to retain the intrinsicality condition on intelligence, perhaps eschewing holistic arrangement. My analysis does not suggest this is wrong, merely that it is inconsistent with the desideratum of decentralized information processing. In general, the propensity of the IMF to guide the intelligent matter research program will be improved by careful consideration of which outcomes can be consistently pursued.

Conclusion

In this chapter I critically assessed Kaspar and colleagues' Intelligent Matter Framework. To do so, I characterized the IMF as having both an *evaluative* and a *guiding* function for research in the emerging field of intelligent matter. To play those roles effectively, the IMF would need to supply appropriate judgements about the intelligence of extant material systems and provide guidance for how future systems should be developed. On this basis, I argued that the IMF's failure to specify any kind of normative criteria with which to evaluate the *behavior* of purportedly intelligent material systems is problematic. Absent a normative criterion, nothing ensures that the relevant functional elements interact in such a way that they implement adaptive behavior, i.e., behavior bringing about desirable consequences. This renders the IMF unsuitable as a guide, since it may lead to degenerate systems, and as an evaluative tool, since it does not track intuitive judgements of intelligence. Thus, I suggested that the IMF ought to be augmented with a normative criterion. I laid out possible approaches, focusing on the distinction between the source of a normative criterion and the way normativity is constituted within system. Ultimately, I suggested that the IMF should introduce a normative criterion which is i) *partially devolved*, that is, requiring certain behavioral conditions to be met, but with the details to be specified on a case-by-case basis according to the goals of specific researchers, ii) *extrinsic,* that is, defined by external agents, as opposed to selected by or emerging from the material system itself, iii) *holistically arranged*, that is, systems ought to robustly behave in accordance with their behavioral norms in virtue of their overall functional organisation, not in virtue of an explicitly represented goal in a central processing location. An IMF augmented with normativity of this kind would be well-suited for guiding and evaluating progress towards the development of intelligent matter.

References

[1] C. Kaspar, B. J. Ravoo, W. G. van der Wiel, S. V. Wegner, and W. H. P. Pernice, 'The rise of intelligent matter', *Nature*, vol. 594, no. 7863, Art. no. 7863, Jun. 2021, doi: 10.1038/s41586-021-03453-y.

[2] M. A. McEvoy and N. Correll, 'Materials that couple sensing, actuation, computation, and communication', *Science*, vol. 347, no. 6228, p. 1261689, Mar. 2015, doi: 10.1126/science.1261689.

[3] A. Walther, 'Viewpoint: From Responsive to Adaptive and Interactive Materials and Materials Systems: A Roadmap', *Advanced Materials*, vol. 32, no. 20, p. 1905111, 2020, doi: 10.1002/adma.201905111.

[4] M. Bauer, 'Normativity Without Artifice', *Philosophical Studies*, vol. 144, no. 2, pp. 239–259, 2009, doi: 10.1007/s11098-008-9208-2.

[5] F. Chollet, 'On the Measure of Intelligence'. arXiv, Nov. 25, 2019. doi: 10.48550/arXiv.1911.01547.

[6] M. Crosby, 'Building Thinking Machines by Solving Animal Cognition Tasks', *Minds & Machines*, vol. 30, no. 4, pp. 589–615, Dec. 2020, doi: 10.1007/s11023-020-09535-6.

[7] J. Hernández-Orallo, 'Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement', *Artif Intell Rev*, vol. 48, no. 3, pp. 397–447, Oct. 2017, doi: 10.1007/s10462-016-9505-7.

[8] E. Nagel, *Teleology Revisited and Other Essays in the Philosophy and History of Science*. New York: Columbia University Press, 1979.

[9] A. Rosenblueth, N. Wiener, and J. Bigelow, 'Behavior, Purpose and Teleology', *Philosophy of Science*, vol. 10, no. 1, pp. 18–24, 1943.

[10] P. Wang, 'On Defining Artificial Intelligence', *Journal of Artificial General Intelligence*, vol. 10, no. 2, pp. 1–37, Jan. 2019, doi: 10.2478/jagi-2019-0002.

[11] P. Godfrey-Smith, 'Environmental complexity and the evolution of cognition', in *The evolution of intelligence*, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2002, pp. 223–249.

[12] P. Lyon, 'The cognitive cell: bacterial behavior reconsidered', *Front Microbiol*, vol. 6, p. 264, Apr. 2015, doi: 10.3389/fmicb.2015.00264.

[13] A. Trewavas, 'Aspects of Plant Intelligence', *Annals of Botany*, vol. 92, no. 1, pp. 1–20, Jul. 2003, doi: 10.1093/aob/mcg101.

[14] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach (4th Edition)*. Pearson, 2020. Accessed: Apr. 25, 2023. [Online]. Available: http://aima.cs.berkeley.edu/

[15] D. Coelho Mollo, 'Intelligent Behaviour', *Erkenn*, May 2022, doi: 10.1007/s10670-022-00552-8.

[16] G. Monreal Santiago, K. Liu, W. R. Browne, and S. Otto, 'Emergence of light-driven protometabolism on recruitment of a photocatalytic cofactor by a self-replicator', *Nat Chem*, vol. 12, no. 7, pp. 603–607, Jul. 2020, doi: 10.1038/s41557-020-0494-4.

[17] S. Legg and M. Hutter, 'A Collection of Definitions of Intelligence', in *Proceedings of the 2007 conference on Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms: Proceedings of the AGI Workshop 2006*, NLD: IOS Press, Jun. 2007, pp. 17–24.

[18] H. A. Simon, 'Rationality as Process and as Product of Thought', *The American Economic Review*, vol. 68, no. 2, pp. 1–16, 1978.

[19] M. Minsky, *The society of mind*, 6. Pb-Pr. in A Touchstone book. New York: Simon and Schuster, 1988.

[20] A. M. Turing, 'Computing Machinery and Intelligence', *Mind*, vol. 59, no. October, pp. 433–60, 1950, doi: 10.1093/mind/LIX.236.433.

[21] J. Hernández-Orallo, 'Twenty Years Beyond the Turing Test: Moving Beyond the Human Judges Too', *Minds & Machines*, vol. 30, no. 4, pp. 533–562, Dec. 2020, doi: 10.1007/s11023-020-09549-0.

[22] R. A. Fisher, *The genetical theory of natural selection*. in The genetical theory of natural selection. Oxford, England: Clarendon Press, 1930, pp. xiv, 272. doi: 10.5962/bhl.title.27468.
[23] S. Okasha, *Agents and Goals in Evolution*. Oxford: Oxford University Press, 2018. doi: 10.1093/oso/9780198815082.001.0001.
[24] R. Carnap, *Logical Foundations of Probability*. Chicago]: Chicago University of Chicago Press, 1950.
[25] S. Legg and M. Hutter, 'Universal Intelligence: A Definition of Machine Intelligence', *Minds Mach.*, vol. 17, no. 4, pp. 391–444, Dec. 2007.
[26] M. Kryven, T. D. Ullman, W. Cowan, and J. B. Tenenbaum, 'Plans or Outcomes: How Do We Attribute Intelligence to Others?', *Cognitive Science*, vol. 45, no. 9, p. e13041, 2021, doi: 10.1111/cogs.13041.
[27] R. J. Sternberg, 'Successful Intelligence in Theory, Research, and Practice', in *The Nature of Human Intelligence*, R. J. Sternberg, Ed., Cambridge: Cambridge University Press, 2018, pp. 308–322. doi: 10.1017/9781316817049.020.
[28] H. L. J. Van der Maas, K.-J. Kan, and D. Borsboom, 'Intelligence Is What the Intelligence Test Measures. Seriously', *Journal of Intelligence*, vol. 2, no. 1, Art. no. 1, Mar. 2014, doi: 10.3390/jintelligence2010012.
[29] V. Braitenberg, *Vehicles: experiments in synthetic psychology*, 9. print. in Bradford book psychology. Cambridge, Mass.: MIT Press, 2004.
[30] R. S. Sutton and A. Barto, *Reinforcement learning: an introduction*, Second edition. in Adaptive computation and machine learning. Cambridge, Massachusetts London, England: The MIT Press, 2020.
[31] J. Yu, B. Wang, X. Du, Q. Wang, and L. Zhang, 'Ultra-extensible ribbon-like magnetic microswarm', *Nat Commun*, vol. 9, no. 1, Art. no. 1, Aug. 2018, doi: 10.1038/s41467-018-05749-6.
[32] X. He *et al.*, 'Synthetic homeostatic materials with chemo-mechano-chemical self-regulation', *Nature*, vol. 487, no. 7406, pp. 214–218, Jul. 2012, doi: 10.1038/nature11223.
[33] I. Rahwan *et al.*, 'Machine behaviour', *Nature*, vol. 568, no. 7753, Art. no. 7753, Apr. 2019, doi: 10.1038/s41586-019-1138-y.
[34] N. Tinbergen, 'On aims and methods of Ethology', *Zeitschrift für Tierpsychologie*, vol. 20, no. 4, pp. 410–433, 1963, doi: 10.1111/j.1439-0310.1963.tb01161.x.
[35] K. Laland, K. Sterelny, J. Odling-Smee, W. Hoppitt, and T. Uller, 'Cause and effect in biology revisited: is Mayr's proximate-ultimate dichotomy still useful?', *Science (New York, N.Y.)*, vol. 334, no. 6062, 2011, Accessed: Jan. 30, 2024. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:1bf0e329-d7dd-4cd0-a354-3bc02b8b4cb0
[36] E. Mayr, 'Proximate and ultimate causations', *Biol Philos*, vol. 8, no. 1, pp. 93–94, Jan. 1993, doi: 10.1007/BF00868508.
[37] K. Neander, 'Functions as Selected Effects: The Conceptual Analyst's Defense', *Philosophy of Science*, vol. 58, no. 2, pp. 168–184, 1991.
[38] G. Ramsey and C. H. Pence, 'Fitness: Philosophical Problems', in *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd, 2013. doi: 10.1002/9780470015902.a0003443.pub2.
[39] E. Jablonka and M. J. Lamb, *Evolution in four dimensions: genetic, epigenetic, behavioral, and symbolic variation in the history of life*. in A Bradford Book. Cambridge (Mass.): The MIT Press, 2005.
[40] K. Neander, 'The Function of Cognition: Godfrey-Smith's Environmental Complexity Thesis', *Biology & Philosophy*, vol. 12, no. 4, pp. 567–580, Oct. 1997, doi: 10.1023/A:1006524203891.
[41] J. Garson, *A Critical Overview of Biological Functions*. in SpringerBriefs in Philosophy. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-32020-5.
[42] J. Garson, 'How to Be a Function Pluralist', *The British Journal for the Philosophy of Science*, vol. 69, no. 4, pp. 1101–1122, Dec. 2018, doi: 10.1093/bjps/axx007.

[43] A. W. Eaton, 'Artifacts and Their Functions', in *The Oxford Handbook of History and Material Culture*, I. Gaskell and S. A. Carter, Eds., Oxford University Press, 2020, p. 0. doi: 10.1093/oxfordhb/9780199341764.013.26.
[44] P. J. Richerson and M. H. Christiansen, Eds., 'The Cultural Evolutionof Technology', in *Cultural Evolution*, The MIT Press, 2013. doi: 10.7551/mitpress/9894.003.0011.
[45] A. Mesoudi and A. Thornton, 'What is cumulative cultural evolution?', *Proceedings of the Royal Society B: Biological Sciences*, vol. 285, no. 1880, p. 20180712, Jun. 2018, doi: 10.1098/rspb.2018.0712.
[46] A. C. Love and W. C. Wimsatt, Eds., *Beyond the Meme: Development and Structure in Cultural Evolution*. University of Minnesota Press, 2019. doi: 10.5749/j.ctvnp0krm.
[47] A. H. Eden, 'Three Paradigms of Computer Science', *Minds and Machines*, vol. 17, no. 2, pp. 135–167, 2007, doi: 10.1007/s11023-007-9060-8.
[48] J. J. P. Tsai and K. Xu, 'A comparative study of formal verification techniques for software architecture specifications', *Annals of Software Engineering*, vol. 10, no. 1, pp. 207–223, Nov. 2000, doi: 10.1023/A:1018960305057.
[49] M. Glinz, 'On Non-Functional Requirements', in *15th IEEE International Requirements Engineering Conference (RE 2007)*, Oct. 2007, pp. 21–26. doi: 10.1109/RE.2007.45.
[50] A. Vaswani *et al.*, 'Attention is All you Need', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Jan. 18, 2022. [Online]. Available: https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
[51] F. Jelinek, R. L. Mercer, L. R. Bahl, and J. K. Baker, 'Perplexity—a measure of the difficulty of speech recognition tasks', *The Journal of the Acoustical Society of America*, vol. 62, no. S1, p. S63, Aug. 2005, doi: 10.1121/1.2016299.
[52] J. Muller, *The Tyranny of Metrics*, NED-New edition. Princeton University Press, 2018. doi: 10.2307/j.ctvc7743t.
[53] D. Manheim and S. Garrabrant, 'Categorizing Variants of Goodhart's Law'. arXiv, Feb. 24, 2019. doi: 10.48550/arXiv.1803.04585.
[54] D. E. Knuth, 'Structured Programming with go to Statements', *ACM Comput. Surv.*, vol. 6, no. 4, pp. 261–301, Dec. 1974, doi: 10.1145/356635.356640.
[55] P. Lanz and D. Mcfarland, 'On representation, goals and cognition', *International Studies in the Philosophy of Science*, vol. 9, no. 2, pp. 121–133, Jul. 1995, doi: 10.1080/02698599508573512.
[56] L. Gugerty, 'Newell and Simon's Logic Theorist: Historical Background and Impact on Cognitive Modeling', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 880–884, Oct. 2006, doi: 10.1177/154193120605000904.
[57] A. Newell, J. C. Shaw, and H. A. Simon, 'Elements of a theory of human problem solving', *Psychological Review*, vol. 65, no. 3, pp. 151–166, 1958, doi: 10.1037/h0048495.
[58] M. Baltieri and C. L. Buckley, 'PID Control as a Process of Active Inference with Linear Generative Models', *Entropy*, vol. 21, no. 3, Art. no. 3, Mar. 2019, doi: 10.3390/e21030257.
[59] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961. doi: 10.1515/9781400874668.
[60] A. Dickinson and B. Balleine, 'Motivational control of goal-directed action', *Animal Learning & Behavior*, vol. 22, no. 1, pp. 1–18, Mar. 1994, doi: 10.3758/BF03199951.
[61] M. Colombo and D. Scarf, 'Are There Differences in "Intelligence" Between Nonhuman Species? The Role of Contextual Variables', *Frontiers in Psychology*, vol. 11, 2020, Accessed: Jan. 30, 2024. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.02072
[62] E. Jablonka and S. Ginsburg, 'Learning and the Evolution of Conscious Agents', *Biosemiotics*, vol. 15, no. 3, pp. 401–437, Dec. 2022, doi: 10.1007/s12304-022-09501-y.
[63] S. Okasha, 'Goal Attributions in Biology: Objective Fact, Anthropomorphic Bias, or Valuable Heuristic?', in *Evolution 'On Purpose'*, P. A. Corning, S. A. Kauffman, D. Noble, J. A. Shapiro, R. I. Vane-Wright, and A. Pross, Eds., The MIT Press, 2023, pp. 237–256. doi: 10.7551/mitpress/14642.003.0016.

[64] D. C. Dennett, *The intentional stance*, 7. printing. in A Bradford book. Cambridge, Mass.: MIT Press, 1998.
[65] E. Nagel, 'Goal-Directed Processes in Biology', *The Journal of Philosophy*, vol. 74, no. 5, pp. 261–279, 1977, doi: 10.2307/2025745.
[66] A. Moreno and A. Etxeberria, 'Agency in natural and artificial systems', *Artif Life*, vol. 11, no. 1–2, pp. 161–175, 2005, doi: 10.1162/1064546053278919.
[67] M. H. Bickhard, 'Process and Emergence: Normative Function and Representation', in *Process Theories: Crossdisciplinary Studies in Dynamic Categories*, J. Seibt, Ed., Dordrecht: Springer Netherlands, 2003, pp. 121–155. doi: 10.1007/978-94-007-1044-3_6.
[68] H. R. Maturana and F. J. Varela, 'Introduction', in *Autopoiesis and Cognition: The Realization of the Living*, H. R. Maturana and F. J. Varela, Eds., in Boston Studies in the Philosophy and History of Science. , Dordrecht: Springer Netherlands, 1980, pp. 5–6. doi: 10.1007/978-94-009-8947-4_1.
[69] K. Ruiz-Mirazo and A. Moreno, 'Autonomy in evolution: from minimal to complex life', *Synthese*, vol. 185, no. 1, pp. 21–52, Mar. 2012, doi: 10.1007/s11229-011-9874-z.
[70] J. Bongard and M. Levin, 'Living Things Are Not (20th Century) Machines: Updating Mechanism Metaphors in Light of the Modern Science of Machine Behavior', *Frontiers in Ecology and Evolution*, vol. 9, p. 147, 2021, doi: 10.3389/fevo.2021.650726.
[71] W. R. Ashby, *An Introduction to Cybernetics*. Springer US, 1956. Accessed: Oct. 26, 2021. [Online]. Available: https://www.springer.com/de/book/9781504128117
[72] H. Von Foerster, *Understanding Understanding*. New York, NY: Springer, 2003. doi: 10.1007/b97451.
[73] M. Dale, J. Miller, and S. Stepney, 'Reservoir Computing as a Model for In-Materio Computing', vol. 22, 2017, pp. 533–571. doi: 10.1007/978-3-319-33924-5_22.
[74] S. A. Benner and A. M. Sismour, 'Synthetic biology', *Nat Rev Genet*, vol. 6, no. 7, Art. no. 7, Jul. 2005, doi: 10.1038/nrg1637.
[75] C. G. Langton, Ed., *Artificial life: an overview*, 5. print. in Complex adaptive systems. Cambridge: MIT Press, 2000.
[76] A. S. Meincke, 'The Metaphysics of Living Consciousness: Metabolism, Agency and Purposiveness', *Biosemiotics*, vol. 16, no. 2, pp. 281–290, Aug. 2023, doi: 10.1007/s12304-023-09531-0.
[77] P. McCorduck, *Machines who think: a personal inquiry into the history and prospects of artificial intelligence*, 25th anniversary update. Boca Raton: CRC Press, Taylor & Francis Group, 2004.
[78] D. R. Hofstadter, *Gödel, Escher, Bach: an eternal golden braid*, 20th-anniversary ed., [Repr.] ed. New York: Basic Books, 1979.