Biology & Philosophy (2023) 38:1
https://doi.org/10.1007/s10539-022-09889-4

# Cognitive science meets the mark of the cognitive: putting the horse before the cart

**Joe Gough**[1,2]

Received: 4 June 2021 / Accepted: 22 November 2022


**Abstract**
Among those living systems, which are cognizers? Among the behaviours of, and causes of behaviour in, living systems, which are cognitive? Such questions sit at the heart of a sophisticated, ongoing debate, of which the recent papers by Corcoran et al. (2020) and Sims and Kiverstein (2021) serve as excellent examples. I argue that despite their virtues, both papers suffer from flawed conceptions of the point of the debate. This leaves their proposals ill-motivated—good answers to the wrong question. Additionally, their proposals are unfit to serve the legitimate roles for characterizations of cognition.

**Keywords** Expected free energy · Mark of the cognitive · Cognition · Cognitive science · Counterfactual cognition

## Introduction

Among those living systems, which are cognizers? Among the behaviours of, and causes of behaviour in, living systems, which are cognitive? Such questions sit at the heart of a sophisticated, ongoing debate (e.g., Adams 2019; Barandiaran and Moreno 2006; Brancazio et al. 2020; Godfrey-Smith 2016a; Lyon 2020; Van Duijn et al. 2006). It is important that 'cognition' be understood correctly in this context. There is a sense of 'cognition', subject to much debate, in which there might be a natural distinction between cognition and perception; similarly, 'cognition' is also used in contrast to emotion. Neither is the sense relevant here, however. 'Cognition' in this context is a notion that includes at least some examples of emotion and perception—indeed for Sims and Kiverstein (2021), affect, given its role in their account of allostasis (their preferred mark of the cognitive), is essential to cognition.

✉ Joe Gough
jg570@sussex.ac.uk

[1] University of Sussex (Philosophy), Brighton, UK

[2] Philosophy, University of Durham, Durham, UK

Published online: 17 December 2022

Two recent, exemplary papers, one by Corcoran and colleagues (Corcoran et al. 2020), and one in response by Sims and Kiverstein (2021) integrate this debate with the free-energy framework, each proposing a mark of the cognitive couched in free-energy theoretic terms. In recent years, there have also been several papers considering the *point* of this debate—considering what is at stake between different characterizations of cognition, why we need a characterization of cognition at all, and what general features we might want a characterization of cognition to have (e.g., Akagi 2018; Allen 2017; Keijzer 2021; Ramsey 2017; Villalobos and Palacios 2021). It is worthwhile to consider the point of the debate not only in order to ensure that we are not wasting our time, but also and moreover because it is vital for enabling us to make well-informed and well-grounded choices between different proposals. I argue that despite their virtues, both proposals suffer from flawed conceptions of the point of the debate. This leaves their proposals ill-motivated—good answers to the wrong question.

Additionally, their proposals are unfit to serve the legitimate roles for characterizations of cognition. Two closely related issues constitute the core of my case: first, they offer characterizations that are undesirably precise (Allen 2017); secondly, they offer 'partisan' characterizations, each taking a strong position on the long-running debate over how widespread cognition is, while I believe that Akagi (2018) is correct that 'ecumenical extensional adequacy' is a key desideratum in a characterization of cognition. Precision and partisanship count against the legitimate roles for characterizations of cognition—summarising the state of the art, highlighting which phenomena are of interest and why, and facilitating intertheoretical and interdisciplinary unification, communication, and collaboration (Akagi 2018; Allen 2017; see also, e.g., Haueis 2021; Neto 2020).

In Sect. Counterfactuals or allostasis?, I lay out the dispute between Sims and Kiverstein and Corcoran and colleagues. In Sect. Proposals from the papers, I argue that their explicit statements about the point of the concept of cognition and of their dispute are insufficient for grounding the debate. In Sect. Cognition and cognitive science, I identify some common ground in the debate at large, that cognition is the domain of cognitive science. In Sect. Counterfactuals and allostasis as target domains, I show that this leaves open various key questions, most importantly, whether or not the concept of cognition serves as a target domain for cognitive science, and suggest that both proposals are best interpreted as suggesting target domains. In Sect. Against prescribing a target domain, I argue that cognitive science has no need of a target domain. In Sect. Targetless characterizations of cognition, I consider legitimate roles for a characterization of cognition, and argue that these count against precision and partisanship. In Sect. Philosophical prescriptions in cognitive science, before concluding, I reflect on the place for philosophical prescriptions in cognitive science.

## Counterfactuals or allostasis?

The recent papers by Corcoran et al. (2020) and Sims and Kiverstein (2021) are among the latest in a sizeable debate about the mark of the cognitive and the nature of cognition. Both papers integrate careful discussion of examples, broader biological and cognitive theorical frameworks, and the aims of cognitive science, in order to reach their conclusions. Both papers are also grounded in the same research paradigm—active inference and the free-energy principle (FEP; e.g., Friston 2012, 2013; Friston et al. 2006; Pezzulo et al. 2015).

Corcoran et al. (2020) argue that the capacity for disengaged, counterfactual cognition, underwritten by a capacity for decoupled representation, and supported by a deep hierarchical model of the environment, is what makes a system a true cognizer. They situate their argument in relation to Godfrey-Smith's (1996) environmental complexity thesis, according to which cognition is fundamentally a tool for dealing with environmental complexity, notably that introduced by the presence of other living systems. They claim that the capacity for counterfactual cognition marks a significant discontinuity in the way systems are able to deal with environmental complexity, and plausibly maps onto Godfrey-Smith's (2002a, b, 2016a, 2016b) proposed distinction between true cognition and mere proto-cognition (where 'proto-cognition' is the name for those ways of dealing with environmental complexity which resemble, but do not count as, cognition).

Sims and Kiverstein (2021) deny that counterfactual cognition is necessary for cognition. They propose instead that a capacity for minimization of *expected* free energy is all that is required for true cognition (they talk variously in terms of 'cognitive behaviour' and 'cognitive causes of behaviour'). Minimization of expected free energy requires selection of action policies that minimize expected future surprise (eg, Friston et al. 2015; Parr and Friston 2019; for further discussion see Millidge et al. 2021). They appear to suggest that minimization of expected free energy is the interesting feature of counterfactual cognition from the perspective of the FEP, and indeed it is minimization of expected free energy that Corcoran et al. emphasise is enabled by counterfactual cognition (e.g., Corcoran et al. 2020, p. 32).

However, Sims and Kiverstein argue for an interpretation of the FEP that does not make strong commitments about the representational apparatus of the described systems, instead claiming that by 'complementing' their environments, self-maintaining systems 'embody' a generative model of that environment. They then argue that on such a construal, minimization of expected free energy is to be found much more widely than anything that can obviously be described as a capacity for counterfactual cognition. In particular, minimization of expected free energy is entailed, they claim, by the kind of prospective, anticipatory action involved in allostasis. Such actions are to be found in systems as simple as single E-Coli bacterium, so they argue.

The second part of their objection to Corcoran et al.'s proposal is that 'cognition' should be understood in a way that is geared towards finding 'gradations in [the] complexity of cognition', and so that cognition '[shades off] into more basic biological process' (Sims and Kiverstein 2021, p. 24). In contrast, they claim that

Corcoran et al.'s proposal, counter to this aim, is geared towards identifying a 'sharp discontinuity' between the genuinely cognitive and proto-cognitive. Defining 'cognition' so that it lines up with such a sharp discontinuity has two disadvantages, they claim: first, it means that apparently cognitive capacities, such as memory and learning, might be found in systems classed as noncognitive by dint of falling the wrong side of the line; secondly, it entails an 'over-intellectualisation of cognition' (ibid.), an idea that they flesh out with an appeal to Morgan's canon (p. 25–26; discussed further below).

Sims and Kiverstein do not deny that Corcoran and colleagues latch onto an interesting kind of (cognitive) system, specifically, one with a deep hierarchical model that enables a capacity for decoupled representation, and hence the kind of disengaged, counterfactual reasoning that we associate with the most impressive instances of human thought (see also Clark and Toribio 1994). In particular, Sims and Kiverstein hold that Corcoran et al.'s proposal identifies cognition with a capacity that is too 'intellectual' to be correctly identified with cognition, marked by too sharp a discontinuity to encourage the search for gradations and shading-off, too exacting to apply to systems that can nevertheless apparently be ascribed such capacities as memory and learning, and not directly related to any FEP-theoretic capacity (although Corcoran et al. claim it is necessary for expected free energy minimization, the bulk of Sims and Kiverstein's argument works towards the denial of this claim). Sims and Kiverstein, as such, propose a capacity to be identified with cognition that is directly lifted from the FEP (expected free energy minimization), shades off into more basic biological capacities, and plausibly applies to all living systems that can be described as learning or remembering (since it plausibly applies to all living systems).

## What's cognition for?

Before I argue against the way the debate currently proceeds, I want to try to find some stable ground by clarifying the *point* of the concept of cognition (for discussion of the points of concepts, see Queloz 2019; Thomasson 2020). Towards the very beginning of this paper, I stressed that the notion of 'cognition' at play here is not the one that gets contrasted with perception or emotion, but the one that includes both perception and emotion. I did not, however, consider what the *point* of this notion of cognition is—what it is for.

First, I argue that the explicit discussions of the point of the concept of cognition offered by Corcoran et al. (2020) and Sims and Kiverstein (2021) are insufficient on their own to tie down the debate (Sect. Proposals from the papers). Next, I argue for a key point of common ground, the link between the domain of cognitive science and the concept of cognition (Sect. Cognition and cognitive science). Finally, I bring out the commonalities and differences between the two proposals under consideration by placing them in a taxonomy of different sorts of view of the link between cognitive science, its domain, and the concept of cognition (Sect. Counterfactuals and allostasis as target domains).

## Proposals from the papers

Both Corcoran et al. (2020, p. 32) and Sims and Kiverstein (2021, p. 24) suggest that the notion is for explaining the relationship between life and cognition (see also Van Duijn et al. 2006), as well as suggesting that it is for distinguishing between cognitive and noncognitive phenomena. However, on their own, these proposals for the point of *cognition* are insufficient.

To say that a concept is for distinguishing those things that fall under it from those that do not seems, at best, trivial, since all concepts with extensions play this role (cf. Cappelen 2018). At first blush, it does not seem to help much to say that the concept of cognition is for explaining how cognition arises from life. Van Duijn et al. (2006) propose that cognition should be identified with sensorimotor control; Corcoran et al. that it should be identified with counterfactual reasoning; Sims and Kiverstein with expected free energy minimization, as indicated by allostasis. None appears to *deny* the existence of the capacity called 'cognition' by the others, nor that the relationship of each proposed capacity to life is an interesting candidate for explanation. It is possible to explain how sensorimotor control, expected free energy minimization, and counterfactual reasoning arise from life, and worthwhile to do so, whether or not any of these capacities is called 'cognition'—and furthermore, labelling any of these capacities as 'cognition' appears to do no explanatory work over and above explaining how these capacities arise from life. These issues might, however, be solved by embedding the proposals in a broader body of theory, or by further specifying what is at stake in distinguishing between the cognitive and the noncognitive.

Both papers do embed their proposals for the point of *cognition* in broader bodies of theory, although for reasons I lay out below, I believe that it is not enough to save either proposal. Corcoran and colleagues appeal to the environmental complexity thesis, and this is an important part of the framing of their paper. Godfrey-Smith (1996) sets up the environmental complexity thesis as a theory about the core adaptive advantage generally conferred by those capacities we count as 'cognitive'. He later weakens the theory somewhat, dropping the idea that it is the 'core' or 'fundamental' advantage conferred (2002a, b). There are two key points here about the way Godfrey-Smith sets up the thesis, both of which are in tension with the way Corcoran et al. mobilize the thesis in their paper.

The first is that Godfrey-Smith is setting up a non-trivial, empirical generalization about the capacities that we call 'cognitive'—*not* stipulatively defining cognition as 'that which is used to deal with environmental complexity' (see especially 2002a; for more on the difference, see Sects. Counterfactuals and allostasis as target domains, Against prescribing a target domain, Targetless characterizations of cognition). Secondly, although Godfrey-Smith insists on a distinction between nongenuine, 'proto-' cognition and genuine cognition (for criticism, see Lyon 2020), he also insists that this boundary is likely to be irredeemably vague, and unhelpful to try to precisify (see especially 2002a). Conversely, Corcoran et al. propose to define cognition such that it is a special way of dealing with environmental complexity, largely to make the distinction between proto- and genuine cognition precise. This is not only in tension with Godfrey-Smith's views, but also undercuts the empirical nature

of the thesis—this is not exactly a fatal flaw, but it does render the appeal to Godfrey-Smith somewhat confusing, and does not clarify what Corcoran and colleagues see as the point of the concept of cognition. Corcoran et al. (2020, p. 32) do express some disagreement with Godfrey-Smith, suggesting that talking of non-cognitive (by their lights) systems as cognitive, or as grading into the cognitive, may 'obscure a fundamental *dis*continuity' (emphasis in the original), but this surely presupposes either that their definition of cognition is correct, or that there can be no significant discontinuities between cognitive systems.

The framing of Sims and Kiverstein's paper centres on an extended appeal to Morgan's canon. In particular, they hold that Morgan's canon and attendant worries about animal psychology place a double burden on theorists, the burden of avoiding underestimating the complexity of seemingly simple systems like bacteria, while also avoiding overintellectualizing their capacities. For them, 'underestimating' a system appears to mean not labelling it as 'cognitive' when it ought to be, and 'overintellectualizing' a cognitive capacity (or 'cognitive achievement'; 2021, p. 25) appears to mean describing its operation in excessively sophisticated terms (e.g., describing E-Coli's anticipatory allostatic behaviour as supported by counterfactual reasoning). This latter worry does not seem directly to speak to whether or not a phenomenon should be labelled cognitive, since it applies only to phenomena already acknowledged as cognitive.

The former worry, that it risks underestimating seemingly simple systems to deny them 'cognitive' status, is more directly relevant. The basic issue is that adopting a more restrictive definition of 'cognition', and thereby denying the cognitive status of, e.g., bacteria, need not 'underestimate' bacteria or their achievements. Say a certain species of bacteria is capable of rudimentary forms of epistemic action. Imagine a theorist who claims that only systems capable of consciously undertaking epistemic actions are cognitive, denies this kind of bacteria consciousness, and therefore denies they are cognitive. This does not mean that the theorist denies 'underestimates' this kind of bacteria—the theorist might fully acknowledge, and be wholeheartedly blown away by, the basic forms of epistemic action that it undertakes. They just might also think there are reasons not to label such behaviour—impressive though it may be—as 'cognitive'. They might, for example, think it is amenable to saliently different models, or that it belongs to a class of interesting phenomena so disparate that cognitive science would dissolve if it were to adopt this class as its subject-matter. Analogously, to deny that an extremely sophisticated robot is 'alive' is not necessarily to deny the impressiveness of its achievements; it might merely reflect a theoretical preference for a notion of life according to which it essentially arises from protracted processes of natural selection.

Julian Kiverstein is generously serving as a reviewer on this paper. He has buttressed this appeal to Morgan's canon by clarifying two worries that lie behind it. One worry is about 'researchers that take human cognition to be the standard of what counts as cognitive relative to which all non-humans fall short.' The other worry is that 'many researchers assume non-cognitive behaviour to be rigid and inflexible whereas this is not the case.' I share the view that both classes of researcher are mistaken.

The first issue, that many researchers taken human cognition as the standard, is a genuine issue in my view, and a genuine problem to be solved, but not a problem that can be solved by a mark of the cognitive (see also Sects. Counterfactuals and allostasis as target domains; Philosophical prescriptions in cognitive science). Sim and Kiverstein's argument *presupposes* that the human case is not the standard, and argues from that presupposition to a characterization of cognition. If it is intended as a refutation of researchers who think that 'cognition' is defined in relation to humans, it fails, because it begs the question against those researchers.

On to the second issue, that many researchers wrongly assume that noncognitive behaviour is rigid and inflexible. One way to understand this worry renders it irrelevant: by this version, there is genuinely noncognitive behaviour that is nonrigid and flexible, and researchers wrongly assume that it is nonrigid and inflexible. This first way of reading the worry seems to me to *undercut* not support Sims and Kiverstein's argument, since one who wishes to deny cognitive status to bacteria can just highlight that there is this oft-neglected category of nonrigid, flexible, yet noncognitive behaviour for bacteria to find a home in. This seems especially true where their opponents, Corcoran et al., are supporters of the free energy principle (which ostensibly identifies a wide domain of flexible, nonrigid capacities and processes), but wish to identify *cognition* with only one small part of this domain. They, surely, do not therefore believe that the living world divides up into the cognitive and the rigid-and-inflexible.

The second way of reading this worry is as identifying a tension to be solved by a liberalized notion of cognition: 'researchers think that all noncognitive behaviour is rigid and inflexible, so let's call all the nonrigid, flexible behaviour cognitive!' I have a lot of sympathy for this position, as it happens, but if this is the argument, I do not think it benefits from being framed as offering a mark of the cognitive (see also Sect. Philosophical prescriptions in cognitive science). Most of the paper, if this is the argument, is rendered confusingly irrelevant.

In sum, I am not convinced that either paper offers sufficient resources to assess their proposals without further constraints. This in itself is not intended as any great criticism of the papers, since I believe that their arguments proceed by presupposing a widely-held view of the point of the concept of cognition.

## Cognition and cognitive science

This brings us to a point of fairly widespread agreement in the debate, which I believe can be used as a fixed point to explore these proposals: *cognition* is for demarcating the domain of cognitive science (Akagi 2018; Allen 2017; Keijzer 2021; Ramsey 2017). There are historical reasons to suspect that the notion of cognition in play is that which is defined in relation to cognitive science—as Boden (2006) points out, prior to the founding of cognitive science, cognition was *defined* to exclude emotion and affect. Cognitive science, as a self-conscious, interdisciplinary exercise, arose in the 1950s, although it grew in part out of the cybernetics of the 1940s. Cognitive science, however, was not yet called 'cognitive science'—much of the work in 1950s went under the simple name 'computer simulation', until

later the term 'cognitive studies' took hold in the early 1960s, before gradually morphing into 'cognitive science' by the mid-1970s (Boden 2006).

There are a few main reasons that the term 'cognitive' took off, according to Boden (2006), based on the account of those primarily responsible. Although the term was, at the time, defined to exclude emotion and affect, no-one wanted to exclude those things from being part of the domain of cognitive science. Instead, they wanted to mark a contrast with behaviourism, and offer a characterization of the new psychology's subject-matter that seemed less trivial and redundant than 'mental'. In the context of the early 1960s, where many cognitive scientists were focussing on cognition (in the narrow sense of perception, language, memory, and problem solving), the term seemed a natural fit (Boden 2006). Through these historical accidents, the term 'cognition' came to be associated with a new concept, one whose *point* is to pick out the subject-matter of cognitive science.

Beer (2021) recently discussed the origin of the phrase 'minimal cognition', a bastardization of his 'minimally cognitive behavior'. When offering a putatively representation-free account of certain organism-level behaviours, Beer found to his frustration that his work was often viewed by cognitive scientists as irrelevant. His work was perceived as irrelevant for reasons best captured by Clark and Toribio (1994)—the worry was that the behaviours he modelled were too importantly disanalogous from, and too simple compared with, paradigmatic, 'genuine' cognition to be relevant to cognitive scientific debates over representation. The phrase 'minimally cognitive behavior' was intended by Beer to get around this problem, and to capture the idea of 'the simplest behaviour deemed worthy of a cognitive scientist's attention'.[1] This vignette contains, I think, a deep truth about the concept of cognition—that it is used, at the most abstract level, to demarcate the domain to which cognitive scientists ought to pay attention.

## Counterfactuals and allostasis as target domains

Even accepting that the concept of cognition is for demarcating the domain of cognitive science, this leaves two key background issues unsolved. The first is this the issue of what might be called the 'direction of fit' (Anscombe 1957; Platts 1979, p. 257). Some proposals are based on the idea that it is up to cognitive science to gradually determine and discover its proper domain,[2] and that the concept of cognition is defined to refer to this to-be-revealed domain whatever it may turn out to be (e.g., Allen 2017; Figdor 2017, 2018; Newen 2017; see also Peirce 1878). I will refer to this class of proposals as 'targetless', and the other class as 'target-driven'. Unlike

[1] This is not to say that I think that Beer's attempt to get around the problem was successful, an issue on which I do not have a strong view for reasons that will become apparent.

[2] I say 'determining/discovering' because although I think that from an internal perspective (see, e.g., Lakatos 1971), the research programmes of cognitive science are best understood as discovering the domain of cognitive science, I also think that from a more sociological perspective, they can be understood as determining our use of 'cognition'.

targetless proposals, target-driven proposals identify a target domain, containing all and only the things that cognitive science ought to study.[3]

Targetless proposals see cognitive science as gradually expanding or shrinking its current remit through the interaction of, at least, the goals of cognitive science, the scope of its models and methods, the original pretheoretic area of interest, the paradigm cognitive capacities that cognitive science originally set out to explain, and perhaps paradigm cognitive systems (normally humans; e.g., Rupert 2013; cf. Figdor 2018; Lyon 2006). Importantly, targetless proposals need not be so flat footed as to claim that anything that the tools of cognitive science can explain forms part of its domain (cf. Ramsey 2017).

Although there are key differences, many targetless views of cognitive science see it as proceeding by working outwards from certain paradigmatically cognitive capacities and systems, incorporating more capacities depending on certain sorts of salient similarity to these paradigmatically cognitive capacities, and incorporating more systems depending on whether they instantiate these capacities (e.g., Allen 2017; Figdor 2017, 2018; Lyon 2015; Newen 2017). Paradigmatically cognitive systems and capacities do not constitute a 'target domain' because it is essential to the way that 'targetless' cognitive science proceeds that this class, the 'paradigmatically cognitive', be used also to identify potential new targets of explanation. 'Working outwards' from the paradigmatically cognitive is guided and heavily informed by amenability to similar models and methods, relevance to the core interests of cognitive science, and other dimensions of salient similarity. A core idea of such views is often that we should let 'the productivity of research programs in cognitive science guide the extension of language to new contexts' (Allen 2017, p. 4240).

A brief tangent is required here. One might worry that acknowledging 'paradigm' cognitive systems, especially if this is set partly in deference to the actual historical remit of early cognitive science, because the question in favour of 'anthropocentric' and against 'biogenic' approaches to cognition. It does not. Importantly, it would not mean humans are 'more' cognitive, but rather that they are more useful in judging whether another system is 'cognitive'. Of course, it is possible to deny that humans are the paradigmatic cognitive systems, and that there are paradigmatic cognitive systems (e.g., Figdor 2018). However, there may be *methodological* justifications for treating humans as the paradigm case, for example, a special interest in explaining human capacities (e.g., Heyes 2014, 2015; Wundt 1907). Even conceding that humans are the paradigm cognitive systems *and* that they have a special place in the *goals* of cognitive science does not guarantee an anthropocentric approach. Indeed, without treating humans as the paradigm case, it is hard to understand many of the key arguments for the biogenic approach. For example, Lyon (see especially 2022) argues that there are 'basal' cases of cognition in extremely simple biological

[3] One major problem in much of the debate so far, in my view, has been the failure to distinguish between these two types of proposal, often leading those offering one type of proposal to see rival proposals as silly or *clearly* wrong. One thing I hope comes out of this section and Sect. Against prescribing a target domain is that I think target-driven proposals, while wrong, are far from silly, and fail in fairly subtle ways. This is partly autobiographical: a previous draft of this paper failed to draw the distinction.

systems by arguing that these basal cases are salient similarity to human cases, especially in being amenable to similar models, and most of all on their relevance in explaining the human case. The relevance of such concerns presupposes and hinges on Lyon treating humans as paradigm cognitive systems in the relevant sense.

Conversely, target-driven proposals are based on the idea that the concept of cognition provides a target at which cognitive science ought to aim. Ramsey (2017, p. 4207) expresses the core idea of such proposals: that cognitive science and cognition should be defined 'in terms of its relevant explananda, in terms of what it is we want explained'. This latter class of proposals faces a second issue. Some are presented as nonrevisionary, and see the concept of cognition as at least roughly the same as the intuitive folk concept of mind: on such a view, cognitive science ought to aim to explain those phenomena that intuitively count as 'mental' or 'psychological'.

For example, Ramsey (2017) articulates a nonrevisionary target-driven proposal according to which cognitive science requires a target domain. In particular, he claims that cognition is best understood 'as a crudely defined cluster of capacities and mental phenomena', and that '[a] theory is a cognitive theory if it helps us to understand a capacity or process or phenomenon that we are pre-disposed to regard as psychological in nature' (p. 4208). Here, Ramsey appears to treat 'mental', 'cognitive', and 'psychological' as synonymous. Another proposal along these lines is offered by Clark (2011).

Other target-driven proposals are presented as revisionary: the intuitive concept of mind is seen as an inappropriate target domain, and a new, more appropriate target domain is offered. These revisionary target-driven proposals have some similarities with targetless proposals: they tend to be informed by trends in cognitive science, views about the possible range of its models, and so on. Even so, they ultimately aim to set a target domain for cognitive science, rather than primarily seeing the proper domain of cognitive science as something to be revealed as cognitive science progresses and matures.

For example, Keijzer (2021) articulates a proposal that like Ramsey's is target-driven, claiming that it is desirable that cognitive science have a 'clear and stable' target domain (p. 137), but differs on the proper target domain. His proposal is offered as revisionary, claiming that initially, the target domain of cognitive science was the mind, or at least 'remained intrinsically bound up with the pre-existing and long-standing notion of the mind' (p. 138). The term 'cognition', he claims, was adopted because it 'provided a scientific, naturalistic phrase that stressed a modern non-dualistic view on the mind that could be articulated in terms of information processing and computation' (ibid.).

However, he thinks that the domain of cognitive science *should* be tied to an 'empirical scientific concept' that can be 'adapted to scientific findings and theorizing' (p. 146). Additionally, he thinks that so long as *cognition* is tied to *mind*, it cannot play this role because '[m]ind is a key concept within our culture that is central for many topics ranging from responsibility, free will, using reasons, being rational, and so on'. He thinks that the intuitive concept of mind therefore frustrates the ability of cognitive science to acquire a stable target domain (see also Clark 2010). His proposal is therefore to set *cognition* free, and untether it from *mind*.

Even so, he proposes a *new* target domain for cognitive science—cognitive science, in Keijzer's view, ought to study all living systems, and in particular, it ought to focus on studying 'cobolism', 'the systematic ways in which each living system encompasses structures, processes and external events that maintain the fundamental metabolic processes that constitute the core of each living system' (2021, p. 137). Rather than approaching the life-cognition boundary by aiming to distinguish between living and cognitive systems, as Corcoran and colleagues do (see Sect. Proposals from the papers), Keijzer's approach is to focus on the distinction between cognition and metabolism as aspects of living systems. This approach is also precedented in the work of Godfrey-Smith (see especially 2016b).

There is another important distinction among target-driven proposals. Strongly target-driven proposals specify a target domain for cognitive science which is also supposed to be its ultimate domain. This tends to be tied to the view that cognitive science is (or at least ought to be) the study of some currently-specifiable natural kind (e.g., Adams 2018).[4] It is this kind of view that Allen (2017, p. 4234) accuses of proceeding by 'definitional fiat', and that Keijzer (2021, p. 147) accuses of 'conceptual stipulation'.

Weakly target-driven proposals give up on the idea that the current target domain of cognitive science should also be presented as the ultimate domain of cognitive science. Instead, target domains are understood as at least somewhat provisional and revisable in light of empirical discoveries. Keijzer sees such target domains as part of '[a] standard scientific bootstrapping process where theorizing and empirical work coevolve' (2021, p. 147).

I believe that Sims and Kiverstein's (2021) proposal is best understood as an elaboration of Keijzer's (2021) position, and hence as a revisionary, weakly target-driven proposal, offering a target domain for cognitive science. The link between Sims and Kiverstein's account, and that of Keijzer, is confirmed by Kiverstein in his role as a reviewer on this paper. Sims and Kiverstein offer a formal elaboration of the nature of Keijzer's 'cobolism', by offering a formal elaboration in free-energy theoretic terms of the nature of allostasis (as minimization of expected free energy), where allostasis is among the most fundamental and most important forms of cobolism (see also their footnote 9).

The view of Corcoran et al. (2020) is not so obviously tied to any of the approaches discussed above. I do not think that it is charitable to interpret their proposal as targetless, largely for reasons I discuss in Sect. Targetless characterizations of cognition. Additionally, I am not sure how one might justify their proposed mark of the cognitive on such a view. The best option I can think of is that one might think that cognitive science will stop at the first major discontinuity (in ways of dealing with environmental complexity) that one reaches as one moves away from what they see as the paradigm cognitive system, humans. According to Corcoran

[4] For various reasons, I minimize my mentions of 'natural kinds' in this paper, partly because according to some accounts, the subject-matter of *any* well-ordered science will by definition be a natural kind (e.g., Spencer 2016), and partly because 'natural kind' is so poorly and variously defined as to confuse more than it clarifies matters (e.g., Hacking 2007).

and colleagues, this is the discontinuity between systems with hierarchical architectures, and systems with architectures that support counterfactuals. Absent a reason that cognitive science ought or is likely to stop at this discontinuity though, such a proposal would be unmotivated.

Taking their proposal as target-driven, I think that it is clearly revisionary. Their definition of cognition is too restrictive to align with any intuitive notion of mind or mentality—disengaged counterfactual cognition is a small part of our 'mental' lives, and describes the activity of very few of our 'mental' capacities. Its closest link to an intuitive notion of mind is to the idea of 'having a mind'. Relatedly, they are particularly interested in demarcating which *systems* are cognitive (see especially the first paragraph of p. 32, and the appeal to Godfrey-Smith therein). Even here, the intuitive notion of *having a mind* does not line up precisely with their technical notion of *being a cognitive system*, since they are willing to deny cognitive status to systems capable of 'learning, memory, and decision-making' (p. 31; this is critiqued by Sims and Kiverstein, p. 25). Even so, one might think (along the lines of Keijzer) that *having a mind* is not a useful scientific notion. One might, on such a view, see Corcoran et al.'s proposal as identifying the scientifically interesting category of systems closest to the 'folk' notion of *having a mind*. It is not clear to me whether their proposal is weakly or strongly target-driven, but I will dismiss both kinds of approach in Sect. Against prescribing a target domain.

In this section, I have tried to find some common ground from which to assess the two proposed marks of the cognitive. In Sect. Proposals from the papers, I argued that neither Sims and Kiverstein, nor Corcoran and colleagues, explicitly offer a satisfactory account of the point of the concept of cognition, and therefore of the stakes of the debate. In Sect. Cognition and cognitive science, I argued that the core point of the concept of cognition is demarcating the domain of cognitive science. In this section, Sect. Counterfactuals and allostasis as target domains, I considered two further background issues (the direction of fit between the domain of cognitive science and the concept of cognition, and the relationship between the concept of cognition and the concept of mind), in order to better flesh out the nature of the two proposed marks of the cognitive. I suggest that both are both understood as revisionary target-driven proposals, trying in an empirically and theoretically informed manner to find relatively stable target domains for cognitive science, severing the link between *cognition* and the intuition-governed folk notion of mind.

It is worth noting that the distinction between targetless, weakly target-driven, and strongly target-driven proposals crosscuts the question of whether there is a mark of the cognitive. Strongly target-driven proposals identify a mark of the cognitive that characterizes *both* the target-domain of cognitive science, *and* the ultimate domain of cognitive science. One way to look at weakly target-driven proposals *and* targetless proposals is as denying that there is a mark of the cognitive because they deny that any characterization should play *both* roles. A more liberal understanding of the 'mark of the cognitive' might identify the mark of the cognitive with whatever characterization fulfils just *one* of these roles. One could then construe the characterizations of provisional target-domains as provisional marks of the cognitive.

Alternately, one could construe the mark of the cognitive as being whatever cognitive-scientific properties demarcate the ultimate domain of cognitive science. For

a supporter of targetless proposals, this is the *only* kind of 'mark' that might exist. Here, there is room for disagreement among proponents of targetless proposals, and among proponents of weakly target-driven proposals: the positions as I have characterized them do not obviously have any entailments regarding the existence of a mark of the cognitive in this sense. They do, however, entail that *if* there is a mark of the cognitive, it *cannot* be known to us presently, since we cannot know the ultimate boundaries of the domain of cognitive science without first answering all the empirical and practical questions that appropriately inform the placement of this boundary. The mark of the cognitive, in this sense, can only follow along behind the practice of cognitive science, it cannot take the lead.

## What's the point of characterizing cognition?

If the above is correct, then the two proposed marks of the cognitive represent two diametrically opposed revisionary target-driven proposals, each couched in free-energy theoretic terms. Sims and Kiverstein (2021) follow Keijzer (2021) in suggesting a broadening of the target domain compared with the folk notion of mind, while Corcoran et al. (2020) suggesting a narrowing of the target domain. Each settles on a theoretically interesting target domain, that ties in interesting ways into the life sciences more generally and especially evolutionary theory.

I may be wrong in this. However, it does not matter to my argument. I prefer to see these proposals as target-driven, suggesting target domains for cognitive science because, if this is their aim, then they have many features that are virtuous in such proposals. However, as I will argue below, this is an illicit aim (Sect. Against prescribing a target domain). Given this, many of the features of these proposals are serious vices in my view (Sect. Targetless characterizations of cognition). It does not matter if I am wrong about the intended direction of fit because even if I am, the proposals have features that are undesirable for targetless proposals. In Sect. Philosophical prescriptions in cognitive science, I clarify that my opposition to target-driven characterizations of cognition is not allied to an opposition towards philosophical prescriptions for cognitive science, before concluding.

### Against prescribing a target domain

One possible role for characterizations of cognition (of which I see 'marks' of cognition as a special case) is to specify the target domain that cognitive scientists ought to study—that is, a *characterization* of cognition may specify the content of a target-driven proposal about the concept of cognition. Such characterizations might reasonably be expected to be clear and precise, and to pick out a category of reasonable scientific and broader theoretical interest.[5] If their goal is to find a suitable,

[5] It is worth noting that some of those who *do* believe in prescribing a target domain to cognitive science do *not* believe that that target domain can be interestingly characterized, since they believe our ascriptions of mind and cognition should be intuitive and may be ineffable (e.g., Clark 2011; Ramsey 2017 is somewhat in this vein).

principled target domain that might be assigned to cognitive science, then I believe that the papers by Sims and Kiverstein and Corcoran and colleagues do about as good a job as possible at this task.[6] Each identifies an interesting category of interrelated phenomena that are closely related to the paradigm cases in cognitive science's remit. However, I do not believe that this task ought to be performed—I do not believe in prescribing cognitive science with a target domain, provisional or not, and so I do not believe that characterizations ought to be used to play this role.

I will first dismiss Ramsey's (2017) argument for prescribing cognitive science a target domain, before offering two brief arguments against doing so. Ramsey's argument is especially significant because it is the basis for Keijzer's claim that 'to get started, a target domain must be chosen' (2021, p. 147; see also p. 139). The argument Ramsey (2017) offers for holding that cognitive science and cognition should be understood in terms of a given domain of target phenomena and capacities in need of explanation is that this is 'the standard way sciences are defined' (p. 4207). He offers the example of geology, which he sees as studying '[roughly] the formation of mountains and rocks and minerals and so on.'

Interestingly, he also mentions chemistry, claiming that it deals with a very different, albeit overlapping, set of phenomena to geology. He does not specify the subject-matter of chemistry. I think he would have a great deal of trouble if he were to try to do so in similar terms. He would, I think, have a similar amount of trouble trying to specify the subject-matter of physics. The problem, compellingly identified by Hempel (1969) in a rather different context, is that the correct, final domain for physics, and its current domain, come significantly apart. The history of physics is littered with disputes about what physical phenomena there are and what phenomena are physical, as well as discoveries of new physical phenomena, and radical changes in our conception of the domain of physics (see Chomsky 2002; Wilson 2006). The same is true of chemistry, especially given its interactions and boundary disputes with physics (Chomsky 2002). Indeed, a major milestone in the maturation of physics was the abandonment of a target-driven view of its domain as the 'material', understood as comprising mechanisms that operated on principles of motion and contact (one might think that cognitive science is undergoing a similar development).

Saliently, psychology has not operated by taking a target domain according to many historians of psychology, instead progressing in a disorderly manner as techniques, interests, and practical goals develop (Danziger 1990, 1997; Leahey 2018; Rose 1985; Smith 1988). Even more worrying for Ramsey's account, it does not appear that even geology functions with a set target domain. As geology progressed over time, it accrued techniques in service of answering certain questions (particularly the origin of the Earth), and its domain apparently shifted when other pressing questions came along which these techniques could help with (for example, how to find valuable minerals, and later oil). Hemeda (2019, p. 2) characterizes geology as 'the study of the character and origin of the Earth, its surface features and internal

[6] If Andrews (2021) is correct, then I am probably wrong about this, since they argue that the FEP is simply not in the business of demarcating categories or domains.

structure' but highlights as advantageous that this characterization has allowed geology the flexibility more recently to consider 'the atmosphere, biosphere and hydrosphere' as (partly) geological phenomena (see also Sect. Targetless characterizations of cognition). Additionally, according to one popular understanding of the history of geology, the Moon and its craters became securely 'geological' phenomena when it was discovered that they were amenable to geological, in particular stratigraphic, analysis (Hemeda 2019). The point here is that even if some sciences are defined with respect to a target domain, this is far from standard practice, and for many mature sciences is simply not the case (see also Allen 2017). Ramsey's argument from standard practice therefore fails.

There are two further reasons not to believe that cognitive science proceeds by targeting a set domain of phenomena. The first, highlighted by Newen (2017) and Miller (2003), is that core 'cognitive' phenomena like human memory, planning, and perception are also studied by other sciences, such as molecular biology, economics, sociology, and the medical sciences. It is not merely that there is a small overlap between the (uncontroversial) domain of cognitive science and the domains of other sciences, as between geology and chemistry. Instead, the domain of cognitive science is almost completely shared with other disciplines, distinguished from cognitive science primarily—*contra* Ramsey—by their *approach* to that domain.[7]

The second reason is that the domain of cognitive science has in fact been hugely unstable, and has expanded through discoveries of salient similarity between phenomena that were at the time uncontroversially part of the domain of the discipline, and those that were not uncontroversially part of its domain (including amenability to similar models and methods, and relevance to some of the practical goals of cognitive science). Consciousness, emotion, affect, allostasis, and the contemporary notion of stress were not uncontroversially part of the domain of cognitive science at its inception. In fact, they were discussed barely if at all. Even so, emotion and consciousness became an uncontroversial part of its domain as the science progressed, the range of models expanded, and these phenomena and their similarities to core cognitive phenomena became better understood (e.g., Akagi 2018; Boden 2006; Clark 2013; Damásio 1994; Hetmański 2018). Affect, allostasis, and stress, although still not entirely uncontroversially part of the domain of cognitive science, are widely discussed within cognitive science, and frequently modelled by cognitive scientists. One needs to offer a compelling argument that it is somehow harmful for cognitive science to proceed this way, if one believes that this way of proceeding has been or has become a mistake—as perhaps Ramsey (2017) and some of those offering highly conservative definitions of cognition (e.g., Adams and Aizawa 2001) do.

I have a third, weaker, argument against characterizing cognition by specifying a prescribed target domain: I agree with Keijzer (2021) that *mind* is an inappropriate target domain for cognitive science, but I see no way of settling the dispute between

[7] I think that the idea of a characteristic approach is compatible with the interdisciplinarity of cognitive science; it means only that this approach must be, among other things, *interdisciplinary*, and therefore dependent on expertise and techniques currently house in different disciplines, especially as these forms of expertise and techniques interact with one another.

revisionary target-driven proposals without undercutting the motivations for offering a target-driven proposal in the first place. So far as I can see, both Sims and Kiverstein (2021) and Corcoran et al. (2020) describe categories of phenomena that could support orderly, interesting sciences. Sims and Kiverstein (2021) argue that Corcoran et al.'s (2020) proposal is 'not unprincipled, [but] nevertheless unwarranted, and certainly not implied by the FEP' (2021, p. 24); I see no reason that Corcoran and colleagues could not say exactly the same of Sims and Kiverstein' proposal. Both go to great lengths to show in a principled way that their proposals are tied to a scientifically and theoretically interesting FEP-theoretic category, but this is not enough to draw a conclusion about what cognitive science ought to study—a question about which the FEP has no direct implications.

The only way that I can imagine the dispute being settled is by considering more directly what it is useful, feasible, and interesting for cognitive science to study given its models, methods, goals, and pretheoretic aims: exactly the sort of concerns that drive targetless accounts of cognition and cognitive science. It is here, if anywhere, that I believe that the FEP has the most direct implications for the concept of cognition and the direction of cognitive science. If nothing else, the FEP provides formal tools that make it feasible for cognitive science to study a broader range of phenomena, because it uses models and tools that are not too alien to cast phenomena like allostasis and homeostasis as interestingly similar to paradigm cases of cognition. However, this is only one consideration among many for determining what it is presently a good idea for cognitive scientists to study.

## Targetless characterizations of cognition

If I am right, and we ought not to be looking for a target domain that can reasonably be prescribed to cognitive science, then this removes one significant possible role for a characterization of cognition. This does not, however, mean that there is no interesting role for a characterization of cognition on the targetless view (see also Akagi 2018; Allen 2017). One possible role for a targetless characterization of cognition, which I raise mainly to dismiss, is to put forward one's best guess about the final subject-matter of cognitive science. The problem with this proposal is that it is, I hope, clear that if targetless proposals are correct and cognitive science leads the way on setting its domain, no-one is in a remotely good position to make such a guess about its ideal, eventual endpoint at the current time.

Characterizations of cognition can be useful without being target-driven and without guesswork. For example, Allen (2017) suggests that characterizations of cognition should play such roles as 'orienting newcomers to phenomena of potential interest', for which they need be neither precise nor exceptionless—he goes through the example of the characterization of cognition as 'adaptive information processing', a characterization as imprecise as 'cognition', and arguably with exceptions,

such as the maladaptive elements of human psychology.[8] Such a characterization helps to highlight the general range of things that cognitive scientists are interested in, and also to highlight why they are interested in those things. The imprecision of this characterization actually *helps* it to play its job. For example, 'adaptive information processing' is imprecise enough that it can be stretched to cover new kinds of case, especially by taking liberal views of 'adaptive' or 'information processing'. This affords more possibilities for creative work that highlights hitherto-overlooked similarities between uncontroversially cognitive capacities and other capacities not (yet) considered cognitive.

There are other, more general reasons that characterizations of cognition benefit from imprecision. In very general terms, cognitive science is interdisciplinary and expansive, and because of this, at risk of disintegration and dissolution if its subsidiary disciplines cease to interact appropriately (as acknowledged by Allen 2017). In light of this, working characterizations of key concepts might also serve to facilitate intertheoretical integration, communication, including communication of different theoretical perspectives, and other 'bridging' roles that form productive links between disciplines in order to resist disintegration. Importantly, many of these roles are in fact *better* played by imprecise concepts (Haueis 2021; Neto 2020). The reason for this is that imprecision gives space for different researchers and disciplines to conceive of their subject-matter in significantly different ways, while still seeing each other as studying 'the same thing' (and therefore worth talking to).

However, to play these roles—conveying the general idea of what cognitive scientists are interested in and why to newcomers, and helping unify the discipline—it is clearly possible for a characterization to be *too* imprecise. If a characterization is too imprecise, it will not be informative, and it may either fail to clearly apply to paradigm cases of cognition, or be so broad as to be stretched to cover cases that are clearly not cases of cognition. This will not serve to orient newcomers, nor help to integrate the discipline. Ideally, then, what we want is a characterization of cognition with just the right amount of imprecision.

Akagi (2018) offers a proposal for how to characterize cognition (albeit, not a characterization of cognition) that can help to solve this problem. Akagi agrees with Allen that characterizations of cognition are of limited use to working cognitive scientists. Instead, Akagi thinks that the main benefits of characterizing cognition are epistemological benefits for others, including philosophers and the public. In particular, Akagi thinks that a characterization of cognition should make explicit current implicit consensus among cognitive scientists about their domain.

This is, of course, difficult in the face of wildly different views of which systems, capacities, and phenomena are cognitive. To preempt this worry, Akagi suggests that characterizations of cognition should be 'ecumenical'—that is, they should capture the dispute, rather than try to gloss over it and take a side. The problem with any 'partisan' proposal that takes a side, in Akagi's view, is that it

[8] Following Neto (2020), I use the term 'imprecise', where Allen (2017) uses the term 'vague', because I wish to remain neutral about whether the kind of imprecision here is properly classed as vagueness (see also Akagi 2018; Haueis 2021).

represents as uncontroversial and established what is in fact highly controversial and not-yet-established.[9]

Instead, Akagi claims, a characterization of cognition should apply exactly as clearly and uncontroversially to any given case as that case is, in fact, a clear and uncontroversial instance of cognition—an ecumenical characterization should apply entirely uncontroversially to a paradigm case of cognition, and highly controversially to a highly controversial case of cognition. It should, in this way, reflect the current state of the art by capturing the nature of the disputes.[10]

It should, I think, be obvious at this stage that the characterizations of cognition offered by Sims and Kiverstein, and Corcoran and colleagues, do not stack up well against the desiderata on targetless characterizations of cognition. Between humans and E. coli, most living systems are highly controversial as instances of cognition, and therefore ought to be part of the penumbra of an imprecise characterization of 'cognition'. Both proposals significantly reduce the penumbra, and decide one way or the other on these controversial cases (they might, in this sense, be understood as precisifying proposals; Fine 1975). They offer proposals that are partisan, and unduly precise, if they are understood as targetless characterization of cognition—although as I have already stated, I think they are better interpreted as target-driven characterizations, and as failing because cognitive science is in no need of such characterizations.

## Philosophical prescriptions in cognitive science

In this closing section, I wish to consider two interrelated objections. The first is that a mark of the cognitive is required for settling disputes that are strictly *internal to* cognitive science, and held among cognitive scientists. The second is that my position is wrongly in tension with or opposed to philosophers offering prescriptions or guidance to cognitive science.

Let us begin with the idea that a mark of the cognitive is required for settling genuine and legitimate disputes within cognitive science. The idea is that there are many disputes internal to cognitive science over whether phenomena are cognitive, such as the question of the boundaries of cognitive systems, and of the potential cognitive status of simple living creatures. Generally, the way such arguments proceed is by showing that present methodological concerns and empirical findings currently *underdetermine* the placement of some boundary between the cognitive and the noncognitive (see especially Varga 2018). They then appeal to a characterization of cognition. The characterizations offered are generally justified by an appeal to philosophical analyses and intuitions (Adams and Aizawa 2010; Aizawa and Adams 2005), *or* to the potential *explanatory* role and other theoretical benefits

[9] For this criticism to apply, of course, the proposal must be understood as targetless rather than target-driven. See also footnote 3.

[10] While Akagi's 2018 paper does not offer a characterization, Akagi (2021) later argues that these desiderata are met by the characterization of cognition as *the sensitive management of an agent's behavior*.

of the category/construct/property identified by the proposed characterization (Corcoran et al. 2020; Sims and Kiverstein 2021).

My proposal is *revisionary* with respect to the current practice of cognitive science in pretty much only one way: I think that this process is wrongheaded, and unable to legitimately settle the disputes. Any appearance of settling the disputes is entirely spurious. Allen (2017) and Akagi (2018) focus on criticising the more 'philosophical' proposals and approaches. Appealing to the explanatory and theoretical benefits also fails because *cognition* is a subject-matter term, that like *psychological*, *chemical, biological, geological*, and *physical*, we should not expect to have any great explanatory role—it is a mistake, owed to the general overemphasis on explanatory terms in historical views and philosophies of science, to try to treat every legitimate scientific concept as playing such an explanatory role (Spencer 2016). To put the issue informally, demarcating a subject-matter is a big and important enough job that we should not overload the concept with further roles, that will inevitably place competing demands on it.[11]

Of course, it is by placing further demands on the concept, whether these are reached by philosophical analysis or scientific-explanatory work, that we get the constraints required to settle the motivating disputes. This is why trying to impose further constraints is tempting: it makes us able to generate something that looks like an answer. However, these further constraints are not actually relevant constraints on the concept of *cognition* as it is used to demarcate the subject-matter of cognitive science.

One reason I am highly suspicious of this process of seeking and imposing further constraints on the concept in order to settle the dispute is that *both* sides of any given dispute are generally *equally* able to justify their position, because there is no principled basis in these disputes regarding *where* to find these constraints. This is why some authors feel entitled find their further constraints in traditional philosophical analyses, others in evolutionary theory, and others in undesirable cultural views of plants. Nothing in the process precludes a *post hoc* grab-bag of principles picked to justify one's already-chosen answer to the dispute in question.

The solution is patience, and a tolerance for uncertainty. Many of these 'disputes' represent a divergence between research programmes with competing commitments and interests. Surely, the thought seems to go, only one of them can be right, and we should try to work out which. The problem is that *we do not know ahead of time which is right*—or even that only one is, since the appearance of competition may turn out to be spurious. We cannot generally determine which is the correct research programme ahead of time, and have to pursue those competing research programmes to settle (and normally also recast) the disputes between them (e.g., Chang 2004, 2012, 2017). These disputes therefore *look* underdetermined by the practicalities of cognitive science and our current empirical and theoretical knowledge because they *are in fact* underdetermined. Adding arbitrary constraints that let us generate precise

[11] My position is arguably revisionary in that the concept has already been assigned some such roles, but I believe that we can identify a *central role* for the concept (Sect. Cognition and cognitive science), and should attempt to ensure that it can play *that* role well (see also Queloz 2019).

'answers' to these disputes is not *settling* these disputes but *obscuring* their existence and their nature.

Of course, some research programmes are unmotivated, illicitly motivated, obviously hopeless, irrelevant to the goals of cognitive science, or deeply impractical. I am not against cognitive scientists or philosophers pointing this out,[12] and I think that often philosophers are the best-placed researchers to do so (Schliesser 2019). Offering a 'mark of the cognitive' is *not* a good way of offering such guidance: this is, in many ways, the most important claim in this paper. Identifying a mark of the cognitive requires prescribing a target domain and/or making a guess about the ultimate future of cognitive science; when arguing about what cognitive science should study, totalizing top-down prescriptions and Oracle-style guesses about the ultimate future of cognitive science are probably irrelevant distractions, and definitely needlessly more complicated than the question of what we ought presently to study.

My basic position on many discussions of the mark of the cognitive is that they are attempts to offer legitimate guidance to cognitive science, but framed in an unhelpful and incorrect manner. Consider, for example, one of the more recent disputes between Clark (2010) and Adams and Aizawa (2010). There, Adams and Aizawa argue that cognitive science should limit itself to what's within the skin on pain of having a subject-matter so broad that the discipline falls apart. Many of their past arguments have been framed in terms of core, and according to them essential, features (in particular, underived intentionality) of what they claim are the true target phenomena of cognitive science. The issue they raise in 2010, however, is a more practical one—it is an attempt to warn cognitive scientists from making what they see as a mistake that might eventually undermine their discipline's very existence. This latter, practical worry is 'laundered' through a dubious philosophical analysis, thus obscuring the practical point by burying it under a needlessly complicated and contentious theoretical edifice.

This more 'practical' understanding of the debate over the boundaries of cognition also suggests a more practical interpretation of Clark and Chalmers' (1998) original argument: that there is no principled reason for cognitive science not to expand its domain beyond the skin, and several potential benefits if it does so. Likewise, Sims and Kiverstein's proposal can perhaps, in a certain light, be understood as claiming that there is no principled reason from the FEP for cognitive science to limit the living systems that it studies to those capable only of explicit counterfactual reasoning.

## Separating prescriptions and characterizations

I have argued that characterizations of cognition, understood as the subject-matter of cognitive science, ought not to aim to specify what cognitive science *should* study—they ought not to try to specify a target domain for cognitive science, including by trying to settle as-yet-unsettled disputes in cognitive science. While Corcoran

[12] Indeed, this is precisely what I take Allen (2017) to be doing.

et al. (2020) and Sims and Kiverstein (2021) offer principled arguments for interesting potential target domains for cognitive science, cognitive science does not need a target domain. Instead, the proper domain of cognitive science will be gradually revealed by the progress of cognitive science.

This does not mean that there is no role for characterizations of cognition: they can play high-level roles in intertheoretic integration, highlighting phenomena of interest, and summarizing the state of the art. For each of these roles, however, they are well-served by being imprecise and nonpartisan: features that the proposals of Corcoran and colleagues and Kiverstein and Sims lack.

Importantly, an insistence on targetless characterizations of cognition is not allied to a blanket ban on offering prescriptions to cognitive science. Instead, it suggests that prescriptions should be more fine-grained, more practical, and often more short term. Prescriptions for cognitive science are simply not best expressed as characterizations of cognition.

**Acknowledgments** With thanks to Andy Clark and my reviewers, Julian Kiverstein and Colin Allen, for invaluable feedback on earlier versions of this piece.

**Funding** N/a.

**Data availabilty** N/a.

**Code availabilty** N/a.

## Declarations

**Conflict of interest** N/a.



## References

Adams F (2018) Cognition wars. Stud History Philos Sci Part A 68:20–30

Adams F, Aizawa K (2001) The bounds of cognition. Philos Psychol 14(1):43–64

Adams F (2019) The elusive extended mind: extended information processing doesn't equal extended mind. Andy Clark and his critics. Oxford University Press, Oxford

Adams F, Aizawa K (2010) Defending the bounds of cognition. In: Menary R (ed) The extended mind. MIT Press, Cambridge, pp 67–80

Aizawa K, Adams F (2005) Defending non-derived content. Philos Psychol 18(6):661–669. https://doi.org/10.1080/09515080500355186

Akagi M (2018) Rethinking the problem of cognition. Synthese 195(8):3547–3570. https://doi.org/10.1007/s11229-017-1383-2

Akagi M (2021) Cognition as the sensitive management of an agent's behavior. Philos Psychol, pp 1–24. https://doi.org/10.1080/09515089.2021.2014802

Allen C (2017) On (not) defining cognition. Synthese 194(11):4233–4249
Andrews M (2021) The math is not the territory: navigating the free energy principle. Biol Philos 36(3):30. https://doi.org/10.1007/s10539-021-09807-0
Anscombe GEM (1957) Intention, vol 57. Harvard University Press, Harvard
Barandiaran X, Moreno A (2006) On what makes certain dynamical systems cognitive: a minimally cognitive organization program. Adapt Behav 14(2):171–185
Beer RD (2021) Some historical context for minimal cognition. Adapt Behav 29(1):89–92. doi:https://doi.org/10.1177/1059712320931595
Boden MA (2006) Mind as machine: a history of cognitive science. Clarendon Press; Oxford University Press, Oxford; New York
Brancazio N, Segundo-Ortin M, McGivern P (2020) Approaching minimal cognition: introduction to the special issue. Adapt Behav 28(6):401–405
Cappelen H (2018) Fixing language: an essay on conceptual engineering. Oxford University Press, Oxford
Chang H (2017) Epistemic iteration and natural kinds: realism and pluralism in taxonomy. Philosophical issues in Psychiatry IV: classification of Psychiatric Illness. Oxford University Press, Oxford, pp 229–245
Chang H (2004) Inventing temperature: measurement and scientific progress. OUP, USA
Chang H (2012) Is water H2O? Evidence, realism and pluralism. Boston Studies in the Philosophy and History of Science
Chomsky N (2002) In: Belletti A, Rizzi L (eds) On nature and language. Cambridge University Press, Cambridge
Clark A (2011) Finding the mind. Philos Stud 152(3):447–461
Clark A, Chalmers D (1998) The extended mind. Analysis 58(1):7–19
Clark A (2010) Memento's revenge: the extended mind extended. In: Menary R (ed) The extended mind. MIT Press, Cambridge, pp 43–66
Clark A, Toribio J (1994) Doing without representing? Synthese 101(3):401–431. https://doi.org/10.1007/BF01063896
Clark A (2013) Mindware: an introduction to the philosophy of Cognitive Science. Oxford University Press, USA
Corcoran AW, Pezzulo G, Hohwy J (2020) From allostatic agents to counterfactual cognisers: active inference, biological regulation, and the origins of cognition. Biol Philos 35(3):32. https://doi.org/10.1007/s10539-020-09746-2
Damásio A (1994) Descartes' error: emotion, reason, and the human brain. Putnam Publishing, New York
Danziger K (1990) Constructing the subject: historical origins of psychological research. Cambridge University Press, Cambridge
Danziger K (1997) Naming the mind: how psychology found its Language. SAGE, London
Van Duijn M, Keijzer F, Franken D (2006) Principles of minimal cognition: casting cognition as sensorimotor coordination. Adapt Behav 14(2):157–170
Figdor C (2017) On the proper domain of psychological predicates. Synthese 194(11):4289–4310
Figdor C (2018) Pieces of mind: the proper domain of psychological predicates. Oxford University Press, Oxford
Fine K (1975) Vagueness, truth and logic. Synthese 30(3–4):265–300
Friston K, Rigoli F, Ognibene D, Mathys C, Fitzgerald T, Pezzulo G (2015) Active inference and epistemic value. Cogn Neurosci 6(4):187–214. https://doi.org/10.1080/17588928.2015.1020053
Friston K (2012) A free energy principle for biological systems. Entropy 14(11):2100–2121. Retrieved from https://www.mdpi.com/1099-4300/14/11/2100
Friston K (2013) Life as we know it. J Roy Soc Interface 10(86)
Godfrey-Smith P (1996) Complexity and the function of mind in Nature. Cambridge University Press, Cambridge
Godfrey-Smith P (2002a) Environmental complexity and the evolution of cognition. The evolution of intelligence. Lawrence Erlbaum Associates Publishers, Mahwah, pp 223–249
Godfrey-Smith P (2016a) Individuality, subjectivity, and minimal cognition. Biol Philos 31(6):775–796
Godfrey-Smith P (2016b) Mind, Matter, and metabolism. J Philos 113(10):481–506
Godfrey-Smith P (2002b) Environmental complexity, signal detection, and the evolution of cognition. Cognit Animal Empirical Theor Perspect Animal Cogn, pp 135–141
Hacking I (2007) Natural kinds: rosy Dawn, Scholastic Twilight. Royal Inst Philos Suppl 61:203–239

Haueis P (2021) A generalized patchwork approach to scientific concepts. Br J Philos Sci. https://doi.org/10.1086/716179
Hemeda S (2019) A foreword from the editor-in-Chief. J Geol Res 1(1)
Hempel C (1969) Reduction: ontological and linguistic facets. In: Morgenbesser WSSPM (ed) Philosophy, science, and method: essays in honor of Ernest Nagel. St Martin's Press, New York
Hetmański M (2018) Archaeology of cognitive science: Michel Foucault's model of the cognitive revolution. Roczniki Filozoficzne 66(3):7–32
Heyes C (2014) Submentalizing: i am not really reading your mind. Perspect Psychol Sci 9:131–143
Heyes C (2015) Animal mindreading: what's the problem? Psychon Bull Rev 22(2):313–327. https://doi.org/10.3758/s13423-014-0704-4
Keijzer F (2021) Demarcating cognition: the cognitive life sciences. Synthese 198(1):137–157. doi:https://doi.org/10.1007/s11229-020-02797-8
Lakatos I (1971) History of Science and its Rational Reconstructions. In: Buck RC, Cohen RS (eds) PSA 1970: In memory of Rudolf Carnap proceedings of the 1970 Biennial Meeting Philosophy of Science Association (pp. 91–136). Springer, Netherlands
Leahey TH (2018) A history of psychology: from antiquity to modernity. Routledge, London
Lyon P (2006) The biogenic approach to cognition. Cogn Process 7(1):11–29. https://doi.org/10.1007/s10339-005-0016-8
Lyon P (2020) Of what is "minimal cognition" the half-baked version? Adapt Behav 28(6):407–424. https://doi.org/10.1177/1059712319871360
Lyon P (2015) The cognitive cell: bacterial behavior reconsidered. Front Microbiol 6. https://doi.org/10.3389/fmicb.2015.00264
Lyon P (2022) The study of the mind needs a Copernican shift in perspective | Aeon Essays. Retrieved from https://aeon.co/essays/the-study-of-the-mind-needs-a-copernican-shift-in-perspective
Miller GA (2003) The cognitive revolution: a historical perspective. Trends Cogn Sci 7(3):141–144. https://doi.org/10.1016/S1364-6613(03)00029-9
Millidge B, Tschantz A, Buckley CL (2021) Whence the expected free energy? Neural Comput 33(2):447–482. https://doi.org/10.1162/neco_a_01354
Neto C (2020) When imprecision is a good thing, or how imprecise concepts facilitate integration in biology. Biol Philos 35(6):1–21
Newen A (2017) What are cognitive processes? An example-based approach. Synthese 194(11):4251–4268. https://doi.org/10.1007/s11229-015-0812-3
Parr T, Friston KJ (2019) Generalised free energy and active inference. Biol Cybern 113(5):495–513. https://doi.org/10.1007/s00422-019-00805-w
Peirce CS (1878) How to make our ideas clear. Popular Sci Monthly 12(Jan):286–302
Pezzulo G, Rigoli F, Friston K (2015) Active inference, homeostatic regulation and adaptive behavioural control. Prog Neurobiol 134:17–35. https://doi.org/10.1016/j.pneurobio.2015.09.001
Platts M (1979) Ways of meaning. Linguist Philos 4(1):141–156
Queloz M (2019) The points of concepts: their types, tensions, and connections. Can J Philos 49(8):1122–1145
Ramsey WM (2017) Must cognition be representational? Synthese 194(11):4197–4214. https://doi.org/10.1007/s11229-014-0644-6
Rose N (1985) The psychological complex: Social Regulation and the psychology of the individual. University of Chicago Press, Chicago
Rupert RD (2013) Memory, natural kinds, and cognitive extension; or, Martians don't remember, and Cognitive Science is not about Cognition. Rev Philos Psychol 4(1):25–47
Schliesser E (2019) Synthetic philosophy. Biol Philos 34(2):19. https://doi.org/10.1007/s10539-019-9673-3
Sims M, Kiverstein J (2021) Is free-energy minimisation the mark of the cognitive? Biol Philos 36(2):1–27
Smith R (1988) Does the history of psychology have a subject? History Human Sci 1(2):147–177
Spencer Q (2016) Genuine kinds and scientific reality. In: Kendig C (ed) Natural kinds and classification in scientific practice. Routledge, London
Thomasson AL (2020) A pragmatic method for normative conceptual work. In: Burgess A, Cappelen H, Plunkett D (eds) Conceptual engineering and conceptual ethics. Oxford University Press, Oxford, pp 435–458
Varga S (2018) Demarcating the Realm of Cognition. J General Philos Sci/Zeitschrift für allgemeine Wissenschaftstheorie 49(3):435–450. Retrieved from http://www.jstor.org/stable/45176490

Villalobos M, Palacios S (2021) Autopoietic theory, enactivism, and their incommensurable marks of the cognitive. Synthese 198(1):71–87. https://doi.org/10.1007/s11229-019-02376-6

Wilson JM (2006) On characterizing the physical. Philos Stud 131(1):61–99. https://doi.org/10.1007/s11098-006-5984-8

Wundt W (1907) Lectures on human and animal psychology (J. E. Creighton & E. B. Titchener, Trans.). Swan Sonnenschein & Co, London