Artificial Intelligence in Medicine (2008) xxx, xxx–xxx

ELSEVIER

ARTIFICIAL INTELLIGENCE IN MEDICINE

http://www.intl.elsevierhealth.com/journals/aiim

Axioms, properties and criteria: Roles for synthesis in the science of consciousness

Robert W. Clowes *, **Anil K. Seth**

Department of Informatics, University of Sussex, Brighton BN1 9QJ, UK

Received 1 May 2008; received in revised form 16 July 2008; accepted 18 July 2008

KEYWORDS
Artificial consciousness;
Phenomenology;
Machine consciousness;
Axioms;
Artificial methodologies in science

Summary Synthetic methods in science can aim at either instantiating a target phenomenon or simulating key mechanisms underlying that phenomenon; 'strong' and 'weak' approaches, respectively. While the former assumes a mature theory, the latter find its value in helping specify such theories. Here, we argue that artificial consciousness is best pursued as a (weak) means of theory development in consciousness science, and not as a (strong) axiom-driven project to build a conscious artefact. As with the other sciences of the artificial (intelligence, life), artificial consciousness can contribute by elaborating the possibilities and limitations of candidate mechanisms, transforming properties into mechanism-based criteria, and as a result potentially unifying apparently distinct properties via new mechanism-based concepts. We illustrate our arguments by discussing both axiom-driven and neurobiologically grounded approaches to artificial consciousness.

1. Introduction and objective

The project of 'artificial consciousness' (AC; or equally, machine consciousness[1]) has over the last 10 years gained substantial momentum and is increasingly prominent within the resurgent science of consciousness. There are two ways to understand the AC project. On the 'strong' view, AC aims to develop conscious artefacts: machines that actually have experiences. This view is implicit in the consensus statement of the seminal 2001 meeting on AC that: "we know of no fundamental law or principle operating in this universe that forbids the existence of subjective feelings in artefacts designed or evolved by humans." [1]. In contrast, on the 'weak' view, AC provides the means to simulate candidate mechanisms to clarify their properties and limitations, without necessarily claiming that the simulations/artefacts are themselves conscious. Following Di Paolo et al. [2], weak AC models can be thought of as 'opaque thought experiments' because they can articulate specific hypotheses through carefully constrained simulation, and because in so doing they allow otherwise impenetrable phenomena to be elucidated.[2] The distinction between weak and strong

* Corresponding author. Tel.: +44 1273 698226; fax: +44 1273 877873.
E-mail address: robertc@sussex.ac.uk (R.W. Clowes).

[1] We favor 'artificial consciousness' in order to emphasize the continuity with artificial life and artificial intelligence.

[2] See also [3].

0933-3657/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.artmed.2008.07.009

versions of AC can be summarized by the following snapshot definitions:

- The goal of strong AC is to produce an artefact that, in virtue of incorporating the sufficient processes responsible for natural consciousness, actually has full-blown consciousness.
- The goal of weak AC is to model functional and mechanistic processes associated with natural consciousness in order to gain insight into these processes, to promote conceptual clarification and development, and also to generate new technologies that benefit from at least some of the functionality that natural consciousness supplies.

The distinction between strong and weak AC is not new [e.g. [4–6]] and has obvious historical precedents in the other sciences of the artificial, intelligence (AI) and life (AL). As with these other sciences, we argue that weak AC will prevail over strong AC in virtue of providing recognizable scientific advances, although we will see that there is a sting in the tail of this assertion.

Strong AC and weak AC invite very different approaches in practice, which turn on interpretations of 'axioms', 'properties', and 'criteria'. Because AC — and indeed the scientific study of consciousness — is still in a so-called pre-paradigmatic stage (i.e., we lack commonly accepted theoretical and conceptual frameworks within which to interpret data and guide experimentation[3]), one approach to strong AC has been to propose a set of axioms as providing ideally the necessary and sufficient conditions for minimal consciousness. However, while top-down, axiomatic approaches can stimulate otherwise quiescent or directionless research areas, they can be weakened to the extent that competing sets of axioms are inconsistent and may be judged to be arbitrary.

In contrast, weak AC does not require axioms but operates instead in terms of properties and criteria. We use these terms in the following uncontroversial manner. A property (of consciousness) is a feature requiring explanation (explananda), whereas a criterion is a specific condition that can be used to decide the admissibility or relevance of data, or to guide construction of simulations and/or artefacts; in other words, a criterion is *operational*. For example, consciousness has the property that it involves a first-person perspective (1PP). This is not yet a criterion because one cannot (yet) identify a 1PP in data, nor is clear how to build a 1PP into a device. Moreover, unlike axioms, properties can be revised as new theoretical insights and new experimental data become available. The concept of a 1PP may well change as we understand more about how the content of conscious mental states is structured by the complex interplay of egocentric and allocentric representations across brains, bodies and environments. Indeed, new experiments on the artificial induction of so-called 'out of body experiences' already show that 1PPs can be experimentally manipulated, somewhat independently of other aspects of consciousness [8].

Our present objective is to clarify the relations among axioms, properties and criteria and to show that weak AC models can contribute to consciousness science by transforming properties into criteria and by showing how otherwise distinct properties can be understood to arise from common neural mechanisms. More generally and in contrast to the axiomatic approach, weak AC encourages a constant recycling of properties and criteria out of which, eventually, mature theory and relevant empirical data can be expected to emerge.

We start with a short critique of strong AC and especially of axiomatic attempts to specify sufficient criteria for a conscious artefact. We then show how weak AC is continuous with computational neuroscience models that test the properties of candidate neural mechanisms of consciousness, noting in each case the problems with interpreting these efforts as strong AC projects. Next, we set forth some specific ways in which weak AC can contribute to the science of consciousness. In particular, we argue that well-specified models can help unify otherwise diverse properties, and can transform these properties into testable and potentially measurable criteria. Finally we argue that through this gradual deepening of understanding of the relationship between the different properties of consciousness, especially between structural properties posited at the phenomenological level of description and properties at functional levels of description, and through the eventual conversion of some of these properties into criteria, the weak approach to AC could deliver much that the strong would promise.

2. Method 1: axiomatic approaches and strong AC

Newton in his *Principia* provides us with an archetype of the operation of the axiomatic approach in natural science. Although the use of axioms were of great importance in the burgeoning "natural philosophy" of the 16th and 17th centuries,[4] Newton's

[3] This notion of pre-paradigmatic science relies on the model of scientific discovery first advanced in [7].

[4] René Descartes similarly formulated axioms which Newton delighted in pointing out were false.

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009

axioms were new in that they encoded 'laws of nature' of great generality and predictive power. The statement of these laws was taken to be axiomatic in that they both encode basic and true assumptions about the world, and that without them no detailed formulation of physical science could be accomplished. These axioms while themselves remaining unexplained underpinned the prediction of the motion of bodies in the world and the explanation of why those bodies moved as they did. Axioms in this sense are central to prediction and explanation in the physical sciences.

It is with this in mind that we should consider the *axiomatic* approach that forms one of the most well known top-down approach currently being developed in AC by Igor Aleksander and his colleagues [9–11]. Aleksander's approach begins with axioms, which in his case are introspectively derived features of consciousness which are held to be minimally necessary and (ideally) jointly sufficient features of any systems to which we would be likely to ascribe consciousness. Aleksander's five axioms (as of the 2007 version of this approach) are *Presence, Imagination, Attention, Volition and Emotion*.[5] These axioms can be considered as elaborations on the basic relation of consciousness which is held to be *Depiction*: perceptual states that 'represent' elements of the world and their location. The axiomatic approach is best thought of a part of the strong AC project because any artefact embodying the stated axioms would in this framework be ascribed consciousness.[6]

There are both general and specific problems with axiomatic approaches. In the latter category, contrary to Aleksander's 'imagination' axiom it seems difficult to exclude the possibility of a conscious organism (or artefact) experiencing itself as being present in a world without having any sense of the world being other than it is, i.e., a being without imagination. In the original paper the imagination axiom is formulated in the following way: "*A* has internal *imaginational states* that *recall* parts of *S* or *fabricate S*-like sensations." (where *A* is an agent and *S* part of its sensory manifold). Yet it is far from clear that we want to deny consciousness to an agent that depicts the world but is not able to fabricate or recall similar conscious sensations. At the least it would need to be shown why imagination is necessary for depiction.[7] This reveals a more general problem with the axioms as stated in that they themselves appear to stand in some need of explanation. But if this is the case, Aleksander's axioms are really quite different from Newton's; for an understanding of consciousness does not seem to flow from them, but rather, ideally, toward them. They seem more like explanatory targets than entities we would expect to do explanatory work.

A further general problem is that the top-down organization of the axiomatic approach undermines claims of sufficiency, leading to a danger of trivial circularity, i.e., if a system is built to instantiate the axioms as stated, then it is said to *be* a conscious system. But simple concordance with a set of axioms is unlikely by itself to convince that consciousness is really present, especially if suspicions remain that the system so built is not treated even by the researchers who build it as conscious (e.g., if the researchers are not compelled to treat their new creation as they would an animal or a baby), in which case there is little else to do besides revise the set of axioms. And yet, this seems to depart from the very *raison d'etre* of the axiomatic approach. Aleksander is careful to state that his own axioms are subject to revision, but then the term axiom seems misleading.

In short, axiomatic approaches currently suffer from inadequately establishing necessity or sufficiency. In logic, an axiom is carefully defined as a proposition not proven but taken to be self-evident, and even in wider usage the term retains the meaning of expressing a truth that can be taken for granted. As Aleksander and Dunmall write "axioms are ways of making formal statements of intuitive beliefs and looking, again formally, at the consequences of such beliefs". However given the current state of consciousness science all such beliefs need to be considered to be provisional and revisable. We argue below that it is the specific advantage of weak (non-axiomatic) approaches to AC that they do not merely work out the consequences of beliefs, intuitive or not, but helps us augment, revise and

[5] There has been some evolution and restatement from the original set of Depiction, Imagination, Attention, Planning and Emotion developed in [10].

[6] At various points in his writings Aleksander argues that the axioms are only jointly necessary rather than sufficient. However in the concluding section of [10] the authors write, after arguing that then current neural simulators embody three of the axioms, that 'given the development or evolution of the remaining two axiomatic mechanisms, what arguments could be used to deny them consciousness?' (p.18). This seems to evoke the sufficiency claim.

[7] It should be noted that a number of other researchers in the field of AC [12–15] do indeed appear to regard the ability to imagine as being necessary for consciousness. It is unclear if this belief appears more widely in the field of consciousness science. Yet at the least, we want to know why it should be necessary to be able to imagine in order to have any sort of consciousness. This needs to be shown or argued for, not assumed [15]. One reading of Edelman and Tononi's dynamic core hypothesis is that imagination, in the sense of a large repertoire of alternative possibilities, is indeed necessary for consciousness (see Section 5 and [38]).

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009

transform our intuitive beliefs.[8] While axiomatic strong AC approaches do not explicitly rule out such transformations, they do not encourage or facilitate them. More generally, by starting with a set of hunches about what is necessary or possibly sufficient for consciousness and in the absence of agreed upon criteria for deciding contrasting claims among researchers, we are in danger of developing a situation where a hundred flowers bloom but none are ever cut.

3. Method 2: structural explanation and weak AC

In contrast to the axiomatic approach to AC, and comprising the majority of work in AC, is the implementation of detailed models based on neurobiological theory. There are a multiplicity of such 'bottom-up' approaches — see [17] for a useful review — here we pick selected examples to show their potential contributions to consciousness science. Although these approaches can be considered bottom-up it should be noted that they do not exclude consideration of high-level of phenomenological properties of consciousness. Rather these are seen as in need of explanation through relation to properties identified at lower levels. Note that bottom-up approaches can be either weak or strong, depending on the claims of the relevant theories.

Both top-down axiomatic and bottom-up approaches require specification of explananda. While in the former approach explananda are given as axioms, in the latter there is a need for independently justifiable sets of properties. As with axioms, there are currently several competing such sets [e.g. [18,19]] and, unsurprisingly, no *a priori* means of selecting among them.[9] However, as we argue, the distinctive feature of weak AC approaches is that properties and criteria are continually recycled during interaction with experimental data and computational modelling; there is no need to assume that the initial property set will remain unchanged. It is indeed a desirable outcome of weak AC that it challenges and/or deepens initial assumptions about properties (see also [6] in this volume).

Most property lists presented in support of general frameworks for consciousness tend to emphasize exhaustively capturing all relevant phenomenology, without concern for whether these properties can be identified in real physical systems. We suggest that the most useful of the current lists, at least for the purposes of weak AC, should have the following features:

(1) They are based on broad agreement across the communities that are currently researching consciousness, i.e. they should if possible respect both phenomenological research on consciousness *and* findings from neuroscience and cognitive psychology.
(2) They pay particular attention to the properties of consciousness at a phenomenological level of description.
(3) They describe not merely properties but *structural* properties.

First, in order to facilitate any sort of agreement on having explained a property of consciousness it is well — though perhaps not essential — that other members of the community agree that it is indeed a property of consciousness, or of systems that support consciousness, or of conscious mental lives. Second, any candidate theory of consciousness (or AC model) must present or embody an explanation that relates to the phenomenological level, otherwise it will be open to the criticism of having nothing to do with consciousness. Third, given the unlikelihood of explaining "the redness of a rose" by simple gesture toward mechanism, it seems useful to concentrate on what we call *structural* properties of consciousness. These can be understood as aspects or dimensions of the way that the world is presented to us; they are presentational properties, aspects or dimensions of *how* the world is given to us. They are also deep properties in the sense that if any were absent from a putative experience we might be inclined to doubt that the putative experience were an experience at all.[10]

One set of properties which has substantial value for the weak AC approach is Metzinger's list of six multi-level 'constraints'[11] (Table 1) (first developed

[8] This evokes the position that Chalmers [16] calls type C materialism, i.e., a solution to the hard problem although possible within the remits of current physics will nevertheless require a shift in our current theoretical framework.

[9] Another approach would be to attempt to taxonomize the types of conscious mental states. An interesting attempt to this by respecting beeper-based sampling can be found in [20].

[10] We leave open for the moment whether the structural properties referred to herein are jointly necessary and/or sufficient for consciousness although it is likely that some of these properties will turn out to be more central than others and that some will turn out to be necessary for only certain types of conscious experiences.

[11] Metzinger calls these constraints, and develops them at multiple levels, we will call them properties for reasons that will become apparent. Metzinger calls them constraints for, assuming representationalism, it can be said that only those representational system that implement the minimal constraints for sufficiency are considered to be conscious.

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009

Table 1 Metzinger's six constraints or properties of consciousness

Constraint	Description
1. Globality	Individual phenomenal contents are always bound into a global situational context, consisting of a conscious world model
2. Presentationality	The temporal immediacy of experience *as such*, that is embedded into a uni-directional flow. Especially the experience of 'nowness' such that every moment includes an immanent immediate past and presages a future
3. Transparency	The unavailability to attention of preconscious processing stages
4. Convolved holism	The structural feature of phenomenal states whereby we experience objects at once as being wholes and not merely sets of features, but at the same time as often being wholes constructed of parts
5. Dynamicity	Our conscious life emerges from integrated psychological moments, which are themselves integrated into the flow of subjective time
6. Perspectivalness	The reference of conscious contents to a subjective first-person perspective. In general, conscious mental life possesses a focus and comes from a point of view

This table lists Metzinger's six overarching constraints on theories of consciousness, which we can understand as properties of consciousness requiring explanation. While Metzinger breaks these constraints down into phenomenal, representational, and functional levels of description, we take them as starting points for a property-criteria dialog mediated by weak AC models [21].

in [18]).[12] Many of the core constraints in Metzinger's list can be seen as structural properties in just the sense introduced above: they are really dimensions of how any given experience is presented, or, to put this another way, they are conditions on a given state or process being considered an experience.

One of the most useful elements of this particular list is that it builds in both many of the findings of the European phenomenological tradition as well as some of the most interesting discoveries of recent consciousness science.[13] Perhaps more importantly, Metzinger offers a way of describing conscious mental states along a number of axes that can give a good starting point for empirical research. There is not the space to fully explain all of Metzinger's six basic properties here but we will describe those we need as we go. Three properties of particular value are *dynamicity, globality* and *integration into a single global coherent state (convolved holism)*, which appear (albeit with differing terminologies) in many phenomenological accounts and which exemplify the notion of a structural property as described above.

While having properties that carve up conscious phenomenality along a number of dimensions is very useful, good scientific theory also requires criteria for deciding the admissibility and relevance of empirical evidence. The difference between criteria and properties is that the former should be testable in practice and/or implementable in models, while the latter should capture deep intuitions about the target phenomenon. Here we argue that weak AC models provide value in translating properties into criteria and in clarifying the relations among properties, leading in repeated cycles of theoretical development, model construction and model analysis to increasingly mature theory. In the next section we will flesh out this argument by considering two different theoretical frameworks, 'global workspace theory' [23] and the 'dynamic core hypothesis' [24] in relation to existing computational models and certain of Metzinger's properties.

4. Material 1: global workspace theory

Baars' 'global workspace' theory (GW [23]) has played a substantial role in the scientific rehabilitation of consciousness over the last two decades and remains one of the most influential among current theoretical approaches to consciousness. GW theory views consciousness as implementing a central resource that can influence and be influenced by otherwise unconscious special-purpose processors. According to GW theory, conscious contents are globally available for diverse cognitive processes including attention, evaluation, memory, and verbal report. The notion of global availability is suggested to explain the association of consciousness with integrative cognitive processes such as decision

[12] There were originally eleven constraints discussed in *Being No-One* [18] which are then limited to six in the *précis* [21]. We will discuss only the properties drawn from the more limited set presented in the 2005 paper as these are most useful for the discussion here.

[13] Metzinger is somewhat skeptical about accepting the *facticity* of the phenomenological tradition, yet nevertheless many of the constraints appear to speak to and derive from this tradition, although see the controversy in [22].

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009

making and action selection. Also, because global availability is necessarily limited to a single stream of content, GW theory may naturally account for the serial nature of conscious experience; considered a central property of at least human conscious mental life; this is a claim we analyze below [25].

GW theory was originally described in terms of a "blackboard" architecture in which separate, quasi-independent processing modules interface with a centralized, globally available resource [23]. Over the last 10 years, a series of computational models have elaborated on this basic idea in different ways and with different degrees of neurobiological fidelity. These models can be considered as weak AC projects, even if in some cases they have not been described as such by their creators.

Most fully developed among GW models is Franklin's IDA [26], which implements a version of the GW model at the computational level as a nine-step processing cycle (see [27]). As of the 2003 version of IDA a number of interesting psychological features are modelled at this level of description including associative memory, episodic memory, action selection and deliberation. The original version of IDA was developed as a naval dispatching system capable of reassigning sailors at the end of a tour of duty, a task noted by Franklin to be one normally performed by conscious humans. Many functions in the IDA model are carried out by specialized processors (e.g., the calculation of costs, job requirements, sailor availability, etc.) which are coordinated by a GW architecture involving 'attention' and 'broadcast' components embedded in an action and perception loop. In the original version perception and action involved the reading from databases and sending and responding to emails and other such events.

More recent models focus less on GW functionality and more on mapping the GW architecture onto a plausible neural substrate. Dehaene, Changeux and colleagues have proposed a so-called "neuronal global workspace" (see Fig. 1 and [28]) in which sensory stimuli mobilize excitatory neurons with long-range corticocortical axons, leading to the emergence of global activity patterns among workspace neurons. Any such global pattern can inhibit alternative activity patterns among workspace neurons, thus preventing the conscious processing of alternative stimuli. Dehaene's model predicts that conscious presence is a nonlinear function of stimulus salience, i.e. a gradual increase in stimulus visibility should be accompanied by a sudden transition of the neuronal workspace into a corresponding activity pattern. Other recent models have been developed by Wallace [29] and Shanahan [30] which emphasize network theoretic aspects and spiking neural dynamics, respectively.

4.1. GW theory and weak AC

The most obvious way in which GW modelling can contribute to the weak AC project is by enhancing the original GW concepts. Most generally, all GW models predict widespread cortical involvement during consciousness as opposed to localized activity for unconscious controls. Such predictions have been extensively verified and place constraints on

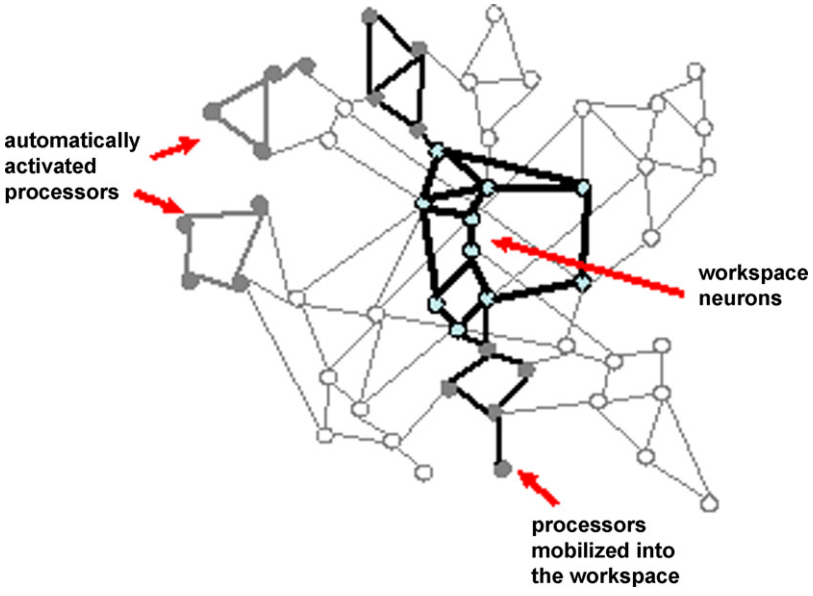

Figure 1 A schematic of the neuronal global workspace. A central global workspace, constituted by long-range corticocortical connections, assimilates other processes according to their salience. Other automatically activated processors do not enter the global workspace. Adapted from [28].

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009

any future theory of consciousness (although see [31]). However, the specific advantages of GW models arise from subtle articulation of observable consequences of the general properties of global access and broadcast when mediated by plausible neural systems. These consequences may not be apparent a priori from armchair consideration of GW theory, hence the appropriateness of thinking of weak AC models as 'opaque thought experiments' (see Section 1).

For example, rephrasing GW theory in terms of a global neuronal workspace (Dehaene) or spiking neuronal dynamics (Shanahan) deepens GW theory and provides opportunities to operationalise otherwise vague concepts. Shanahan's spiking models show how different activity patterns can come to dominate the GW and influence neuronal processes that would be otherwise independent, thus sharpening and operationalising the concept of 'broadcast' within GW theory. As noted above, Dehaene's model predicts that new GW states will arise via fast nonlinear transitions — 'ignition' events — suggesting that particular conscious contents are 'all or none' rather than graded. This is another example of a property transformed into a criterion: experimental data can be analyzed for evidence of such ignition events.

4.2. GW theory and properties of consciousness

GW models can not only articulate new relations between properties and criteria, they can shed new light on the properties themselves. This is perhaps their most valuable contribution to consciousness science.

GW theory can be partly explicated as linking together two of Metzinger's properties (see Section 3.2) and predicting a third. Those it links are, *global availability*[14] the idea that the same content stream can be addressed and is available to different brain processes,[15] and *adaptivity*, i.e. that conscious contents are always oriented toward the flexibility of behavioural control. Under GW theory the adaptivity constraint is operationalised through the notion that neural systems are always undergoing an ongoing competition for scarce resources, because of need for access to the global workspace and thus domination of the control of behaviour. These two properties together predict that a system that implements a global workspace would also implement at least some — but not all — of the functional correlates of the temporal structure of consciousness. The most important property that seems (in part) to come for free from the assumption of a GW architecture is called by Metzinger *dynamicity*.

Dynamicity refers to the typical aspect or attribute of conscious mental states whereby they are not static but are embedded in time in a variety of ways. Conscious mental contents proceed in the serial way often referred to as the 'stream' of consciousness ([25] et seq). Franklin's IDA implementation of a GW architecture [33] nicely illustrates that the interaction of the properties of *adaptivity* and *global availability* when operationalised at functional/computational level of a system can explain something of the seriality of conscious mental states. Given this information processing story it shows why single global states evolve and are conditioned by the need to respect and arbitrate among a number of goals embedded within the system architecture. But the IDA architecture also helps illustrate why GW theory, at least when it is not elaborated at a neural level of description, does not necessarily account for the full complexity of the temporal structure of conscious experience[16]: seriality and fleetingness alone do not a stream of consciousness make.

To see this we should note that *dynamicity* is in part an elaboration of a more basic phenomenological property, what Metzinger calls the *presentationality* constraint, i.e. that every conscious content is anchored in or is a presentation of 'the now'. Consider that every moment of consciousness implies and contains both what has gone immediately before and what will follow immediately afterwards [34]. This is to say that consciousness has a fine-grained temporal structure that was first described in detail by the phenomenologist Husserl [35]. According to Husserl's analysis the world is not presented to us as a bare time-slice but in addition to the 'primal impression' there is both an anticipation of the future or 'protension' and echo of the immediate past or 'retension'. These factors — not to be understood as memory but imminent in experience itself — are often explained in relationship to listening to a tune. The musical note that I am listening to at this precise moment only makes sense

[14] Of Metzinger's original eleven constraints nine are essentially driven by top-down phenomenological concerns and two (not previously discussed) are primarily specified primarily at a functional or neurological level. Global availability is one of these latter two and the other is Adaptivity.

[15] The global availability constraint is strongly linked to and is really a successor concept to Block's notion of functional consciousness [32].

[16] The IDA model makes much use of what are called 'codelets' which are relatively insulated computational mechanisms which run in their own independent processing space, while these are argued to instantiate certain mechanisms which are instantiated neurally in animals they are not themselves a neural instantiation.

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009

and has the character it does have in virtue of what notes have been played before. The content of the experience is thus constrained to have the character it does in part in virtue of a structural property, i.e. the co-presence of the immediate past and an imminent future. Similarly, it is because anticipation is implicit in my experience of the present moment that I can notice a dissonant or mistaken note even if I do not recognise the tune I am hearing. Thus, it is not merely that there is a stream of consciousness but that at each moment the stream is indexed to a 'now' that implicitly contains the immediate past, present, and future: 'the now' has structure. Can GW theory explain this?

Given the IDA implementation of GW theory there is no obvious reason why the currently occurrent 'functionally' conscious mental state should be considered to give rise to this structured present that we have followed Metzinger in calling presentationality. Indeed it seems entirely mysterious as to why there should be a now at all. It should be noted that the temporal structure of the now is actually embedded in a dynamic, continuously changing and sensible temporal evolution (i.e. dynamicity). Thus consciousness is not only continuous but continuously changing; we experience not only a succession of conscious contents but we experience as well the succession itself. However even this more basic feature of the dynamic structure of consciousness — beyond mere succession — is not obviously predicted by the IDA model. Insofar as we follow the strong conception of AC and take dynamicity and presentationality to be proper explanatory targets, then these models seem a failure.

But, things are different from the weak AC perspective. It is important to see that although the IDA model does not obviously predict structural properties of time apprehension this is not straightforwardly a failure of the model, or GW theory, but rather a way in which the model demonstrates a needed enrichment of the theory. Whether this enrichment can take place by developing explanatory resources already present in the theory or else by suggesting new explanatory principles presently unclear.[17] However the methodological point remains. Weak AC models can function as a spur to the further enrichment of given theories even where they 'fail'. Failure in fact can suggest ways that a future research programme might be expanded.

4.3. GW theory and the explanation of structure

The above examples show that weak AC models of GW theory can add value to consciousness theories and generate empirical predictions. Yet it is exceedingly unlikely that we would attribute actual phenomenal consciousness to models such as Franklin's IDA or Dehaene's neuronal global workspace.[18] One reason for this is that we may suspect, implicitly or explicitly, that a functional architecture such as instantiated in IDA should have certain other central structural features that we ascribe to conscious mental systems.

One such structural feature is *perspectivalness*, i.e. the property of our conscious mental life in that it, at least under normal configurations, appears to be tied to a single spatial viewpoint.[19] It does not seem at all likely that a system such as that implemented in IDA whose only sensorimotor mechanisms are the ability to read and write from a database, and to read and write from emails, could possess the minimal sort of spatialisation that would be sufficient to count as a point of view. Nor does analysis of IDA in practice do anything to dispel this suspicion. This usefully points to a theoretical challenge for future versions of GW theory: to explain why a system implementing a GW should be seen as having a point of view.

Challenges such as this are not impossible to meet. Indeed, recent modelling work by Shanahan [37] has focused on embedding a GW architecture in a simulated body and environment, and augmenting it with the capacity to model the sensory consequences of actions taken. Such a combination of embodiment and forward modelling may go some way to explaining the genesis of a point-of-view and thus deepen the explanatory power of GW theory more generally.

In the next section we address some of the same issues but with reference to a very different theoretical framework, provided by the 'dynamical core' hypothesis originated in 1998 by Edelman and Tononi [24].

5. Material 2: the dynamic core hypothesis

The dynamic core hypothesis (DCH [24,38][20]) was developed in the context of the theory of neuronal group selection (TNGS, also known as neural

[17] Following Lakatos [36] it seems likely that a complex research cannot be judged to have failed simply by seeing that its early formulations it cannot predict and explain all the features of the explananda. This can only be judged by the fecundity of the research programme over time.

[18] And indeed Franklin does not make this claim and rather claims that IDA is in Block's [32] sense, functionally conscious.

[19] There are several conscious states which may not in fact involve perspectivalness, for example deep meditative states.

[20] For a review see [39].

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009

Darwinism), a selectionist theory of brain development and brain function [34,40–42]. This theory focuses on a very important and previously overlooked structural feature of consciousness; namely that every conscious scene is both integrated and differentiated [24]. That is, every conscious scene is experienced 'all of a piece', as unified, yet every conscious scene is also composed of many different parts and is therefore one among a vast repertoire of possible experiences: *When you have a particular experience, you are distinguishing it from an enormous number of alternative possibilities*. On this view, conscious scenes reflect informative discriminations in a very high dimensional space where the dimensions reflect all the various modalities that comprise a conscious experience: sounds, smells, body signals, thoughts, emotions, and so forth (Fig. 2). [In terms of Metzinger's list of properties, the simultaneous existence of integration and differentiation may be equivalent to the property of *convolved holism* (see Table 1).] A central claim of the DCH is that conscious qualia *are* the above discriminations [41]. The DCH proposes that the neural mechanisms underlying consciousness consist of a functional cluster in the thalamocortical system, within which so-called reentrant (i.e. massively parallel and reciprocal) neuronal interactions yield a succession of differentiated yet unitary states. The boundaries of the dynamic core are suggested to shift over time, with some neuronal groups leaving and others being incorporated.

Although the DCH and GW theory share an emphasis on global integration, they differ in significant ways. While GW theory stresses functional properties of cortical competition and broadcast in order to predict and explain the phenomenological features of the seriality and fleetingness of conscious mental states, at the heart of the DCH is a different intuition about the structure of conscious phenomenology: that of the simultaneous integration and differentiation of conscious contents.

5.1. The DCH and weak AC

Although detailed neuronal models of the dynamic core are lacking, a notable feature of the DCH is the proposal of a quantitative measure of neural complexity [39,43], high values of which are suggested to accompany consciousness. Neural complexity uses information theory to measure the extent to which the dynamics of a neural system are both integrated and differentiated [43] (Fig. 3). A series of modelling studies have suggested that the distinctive reentrant anatomy of the thalamocortical system is ideally suited to producing dynamics of high neural complexity [44].

The definition of neural complexity marks a major advance for theories of consciousness because it operationalises an apparently contradictory concept: the simultaneous coexistence of dynamical segregation and integration in a single system. This is an excellent example of the transformation of a property (consciousness as discrimination, convolved holism) into a criterion that can be assessed in empirical data and explored in synthetic models [44]. Furthermore, the introduction of one measure

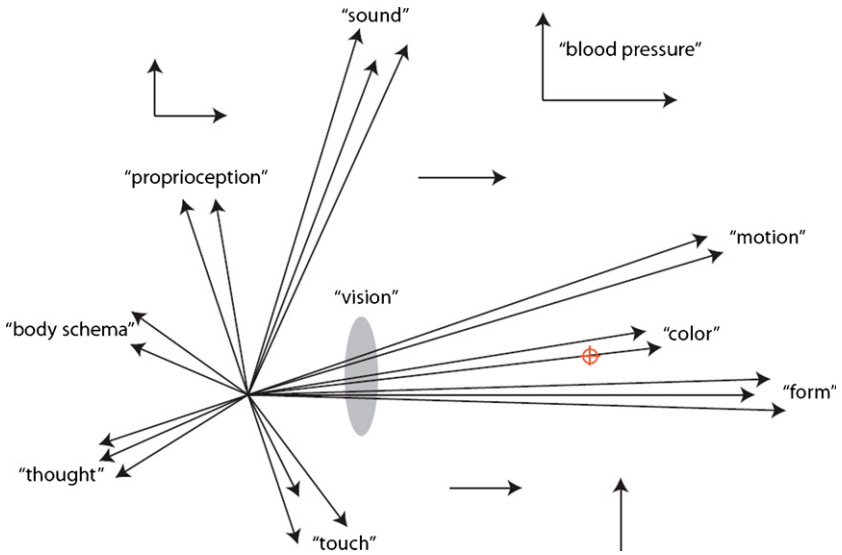

Figure 2 The figure shows an *N*-dimensional neural space corresponding to the dynamic core. *N* is the number of neuronal groups that, at any time, are part of the core, where *N* is normally very large (much larger than is plotted). The appropriate neural reference space for the conscious experience of 'pure red' would correspond to a discriminable point in the space (marked by the small cross). Focal cortical damage can delete specific dimensions from this space.

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009

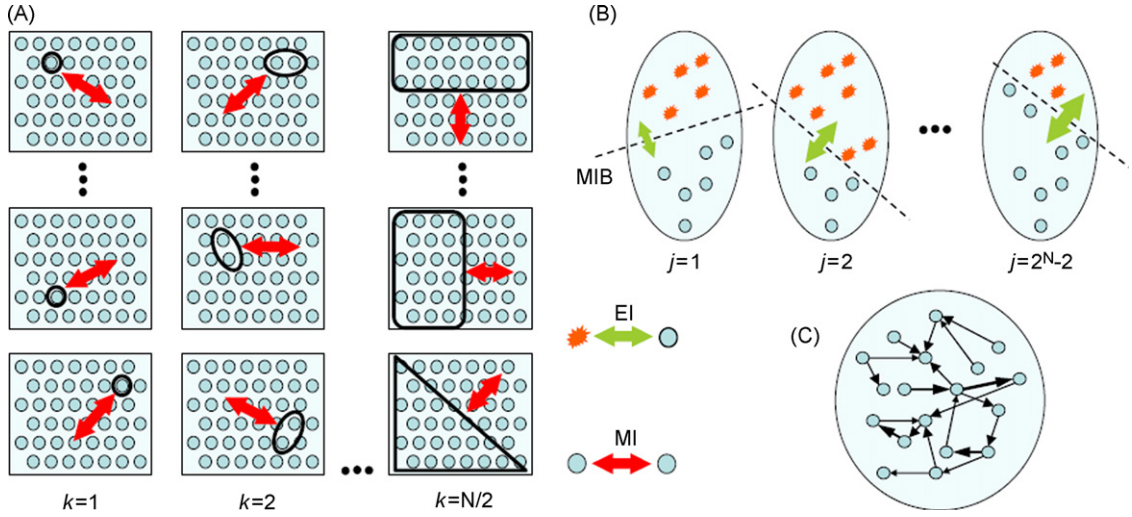

Figure 3 Neural complexity, information integration, and causal density. (A) Neural complexity (C_N) is calculated as the sum of the average mutual information (MI) over $N/2$ sets of bipartitions indexed by k (e.g., for $k = 1$ an average MI is calculated over N bipartitions). (B) Information integration Φ is calculated as the effective information (EI) across the 'minimum information bipartition' (MIB). To calculate EI for a given bipartition (indexed by j), one subset is injected with maximally entropic activity (orange stars) and MI across the partition is measured. (C) Causal density c_d is calculated as the fraction of interactions that are causally significant according to Granger causality. Reprinted with permission from [55].

of complexity naturally encourages its critical analysis and the proposal of alternatives which either serve to correct apparent deficiencies or which aim to operationalise subtly different aspects of the same overarching property.

For example, a criticism of neural complexity is that it is not sensitive to causal interactions in neural dynamics, because it based on the symmetric quantity of mutual information. Yet the brain is fundamentally a causal engine in which neurons physically cause one another to fire. Recently, Seth ([19] et seq.) proposed a measure called 'causal density' which, as implied by the name, measures the overall density of causal interaction generated by a system (Fig. 3).[21] Like neural complexity, this measure is low for a system of independent elements and is also low for a system composed of highly integrated elements. It is maximised somewhere in between, which is the realm of complexity. Unlike neural complexity, causal density is based on a statistical interpretation of causality, i.e. Granger causality [45,46] and is therefore sensitive to causal interactions. (There are other differences, but we will not discuss them here.)

A third measure, 'information integration' [or Φ (phi)] is defined as the amount of causally effective information that can be integrated within a system (Fig. 3 and [47,48]). Like causal density, Φ is sensitive to causal interactions, but unlike both causal density and neural complexity, its value depends only on the dynamics across one partition of a system, the so-called 'minimum information bipartition', which can be thought of as a kind of informational 'weakest link'. Because of this, Φ is a measure of the *capacity* of a system for information integration, rather than the information actually integrated by the system.

In the present context, Φ is distinctive because the theory in which it plays a central role — the 'information integration theory of consciousness' (IITC) — states the very strong position that high Φ is sufficient for consciousness. Therefore, a computational model exhibiting high Φ would not only be a model of consciousness but on the IITC would actually be conscious. This hypothesis is a clear instance of strong AC, and therefore invites the challenge of designing systems that exhibit arbitrarily high Φ while satisfying few if any other intuitions regarding consciousness. For example, simple fully connected Hopfield-type networks seem able to generate arbitrarily high Φ and therefore according to the IITC would be arbitrarily conscious [49].

This aside, the overriding positive conclusion to be drawn is that one proposal for transforming a property into a criterion can initiate a succession of competing alternatives. This population of candidate criteria can then be expected to evolve further under the constraints of (i) ease of application to

[21] Interestingly, Metzinger also uses the term 'causal density' in reference to the constraint of convolved holism [21]. This usage was derived independently of the work by Seth.

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009

empirical data, (ii) match to theoretical prediction when applied to empirical data, and (iii) theoretical insight provided.

5.2. The DCH and properties of consciousness

In Section 4.2 we saw that GW theory, while accounting for seriality in terms of global access and adaptivity, did not obviously explain fully the phenomenology of *presentationality*. Here we assess whether the DCH offers any advantage and our conclusion is that it does, though models illustrating this are still lacking, as are specific measures such as those described above.

Complexity understood as a balance between integration and segregation not only exists in the spatial domain (where it is measured by neural complexity, causal density, and/or Φ); it also exists in the temporal domain, where it has been given the name *metastability* [50]. In direct analogy with neural complexity (though the concepts were developed independently), metastability refers to the property of a certain sort of dynamical system that expresses simultaneously the tendency to settle into a particular dynamical regime and to leave that regime. A metastable system is marked by transients between pseudo-stable states. Although an explicit measure of metastability does not yet exist, it is surely conceivable as a temporal analog of information-theoretic neural complexity and indeed some initial attempts at formalizing this notion have already been made [49].

Metastability may provide a means to operationalise presentationality, to turn it into a criterion or criteria. A system which has states that are partly constituted by tendencies towards new states (that are partly constituted by such tendencies themselves) as well as by the residues of previous metastable states (which themselves incorporate such residues), can be understood as a system in which the 'now' is a concoction of past, present, and future. Having a measure of metastability therefore potentially transforms the property of presentationality into a criterion, to be assessed in empirical data and modelled in simulations.

Even in the absence of a specific measure of metastability, weak AC models can shed light on this issue. For example, recent years have seen impressive progress in large-scale modelling of the mammalian thalamocortical system, culminating in the recent construction of a thalamocortical model containing about one million neurons and half a billion synapses, organized according to detailed multiscale neuroanatomy [51]. Although this model has not yet been analyzed for its neural complexity, and bearing in mind that such analyses will be computationally highly challenging, having such models at least opens the possibility of understanding whether a system organized to generate high values of neural complexity will naturally also generate highly metastable dynamics, or whether such aspects of complex dynamics can manifest separately. If the former, presentationality could be seen as a property that is coherent with convolved holism (simultaneous differentiation and integration) and that arises from a common underlying mechanism. If not, then weight is given to the alternative that other mechanisms may be needed, such as explicit mechanisms for forward/predictive modelling.

We have focused on the example of presentationality not only because it is fundamental structural property of consciousness not accounted for by GW theory, but also because it shows how weak AC can lead to further development of consciousness theory itself. As standardly stated, the DCH and its associated operational measures (neural complexity, causal density, perhaps Φ) have to do with convolved holism, and do not by themselves address or account for presentationality. However, our discussion has suggested how the DCH can be naturally extended to account for this property, in three mutually reinforcing ways: (i) by proper appreciation of the phenomenology of presentationality, (ii) by development of specific measures of metastability, and (iii) by construction of detailed models in which the relation between metastability and complexity can be observed.

6. Discussion

Weak approaches to AC provide a powerful means of deepening our current theories of consciousness. By concentrating on *structural* aspects of conscious mental life rather than qualia *tout courte* they avail of the special mode of explanation that is central to science, i.e. mechanism-based understanding. Further, weak approaches to AC provide a way of operationalising deep intuitions about consciousness and exploring these intuitions through synthesis.

Weak AC theories are of particular use when they allow us to propose ways of relating different structural properties of consciousness through mechanistic implementations. Building models and artefacts presents back to us our theories in a form that is both concrete and complex. *Concrete*, because we can record all aspects of system, rerun it under the same and/or different conditions, and witness in arbitrary detail the impact of specific perturbations; a model or artefact can be subjected to much greater scrutiny than can its natural coun-

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009

terpart. *Complex*, because the models and artefacts of weak AC are normally far from transparent; their full understanding requires careful analysis in much the same way as do natural systems. Hence the useful concept of an 'opaque thought experiment'.

Thus a weak a AC approach, allied to some of the best of contemporary theories of consciousness, and guided by the need to explain structural features of consciousness can both deepen and extend theoretical frameworks. For example, describing in more detail what constitutes *a point of view* such that it might be modelled requires the gradual elaboration of that property of conscious systems and indeed is likely to imply its connection to other properties of that system. Similar progress might also be suggested arising from the principle of dynamicity as described above (Section 3).

We mentioned at the outset that there was a sting in the tail of our assertion that weak AC will prevail over strong. Where is this sting? It lies in the suspicion that future progression in weak AC may inevitably lead toward a strong AC. As one progressively builds in new constraints to the match objections that become apparent through the building of models, so, the models in question may actually tend toward the instantiation of systems that might genuinely be considered conscious. Let us take one more concrete example related to the GW theory.

One might propose for any system to be actually conscious is should be considered to have a point of view.[22] As we have discussed, a limitation of GW theory is that there is no obvious reason to think that the mere instantiation of a GW would lead a mechanical artefact to be ascribed anything like having a point of view or perspective. Once again this is made obvious by the instantiation of Franklin's IDA; it is difficult to see any reason why IDA should be ascribed a point of view and indeed Franklin makes no such claims. Having a point of view, a perspective, seems to imply that an agent need be in some sense spatially located and capable of movement of its perceptual apparatus within that spatial location. Having a point of view requires therefore that an artefact is *embodied,* with sensorimotor abilities that go some way beyond the reading of emails and broadcasting of billeting instructions.

Noticing this, it becomes possible to extend GW models in the right direction, for example by implementing them in a perception—action cycle along with simulated bodies and environments [37]. Of course, other fundamental properties may still be missing; convolved holism, for example, may require a dynamic-core style theoretical approach. It is in this way that the weak approach to AC may be seen as leading toward increasingly stronger statements of AC, for if we can establish the mechanistic basis for each generally accepted structural feature of consciousness then it is at least likely there will be no further limitations on building an 'actually conscious' entity.

There is no particular mystery to this possibility. In the field of artificial life (AL), by analogy, it is increasingly accepted that computer-based models are precisely that: *models* of life (i.e. weak AL) rather than life itself (strong AL). Yet there now exists another overlapping field called 'synthetic biology' in which researchers create new life forms by the artificial synthesis of genetic material and subsequent implantation of that material into surrogate embryos [53]; the consensus here is that these new organisms are in fact alive and are not mere models. It may turn out that an AC model that is sufficiently rich to fully account for all structural properties of consciousness will not to be implementable in computers and will instead require implementation in neural, or some other, material.

7. Conclusion

Over the history of biological science, major advances have been made not only by direct analysis of target phenomena but also by the creation and analysis of artefacts and simulations. From the automata of Jacques de Vaucanson to contemporary neurorobotic devices, insights provided by artefacts have repeatedly exemplified Braitenberg's law of "uphill analysis versus downhill synthesis", the idea that complex phenomena that resist direct analysis can best be understood by analysis of less complex alternatives instantiated in simulation [54]. Artificial consciousness looks set to continue this productive tradition, and as with its cousins artificial life and artificial intelligence, the most rapid intellectual gains are likely to made by embracing a weak version of AC as a necessary step on the way to a strong AC. Whether weak AC will ever be sufficient for strong AC is a question that for now remains open.

[22] Something central to many conceptions of consciousness is that conscious beings are things that it is something it like to be [52]. It is difficult to see how there could be something it is like to be a thing without that thing having in some sense a point of view.

References

[1] Koch C. Final report of the workshop Can a machine be conscious. http://www.swartzneuro.org/abstracts/2001_summary.asp; 2001 (accessed 15.07.08).

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009

[2] Di Paolo EA, Noble J, Bullock S. Simulation models as opaque thought experiments. In: Bedau MA, McCaskill JS, Packard NH, Rasmussen S, editors. Artificial Life VII: Proceedings of the seventh international conference on artificial life. 2000. p. 497–506.
[3] Sloman A. The computer revolution in philosophy: philosophy science and models of mind. Atlantic Highlands, NJ: Harvester Press, Hassocks, Sussex, UK & Humanities Press; 1978.
[4] Holland O. Editorial introduction. Journal of Consciousness Studies 2003;10(4/5):1–6.
[5] Torrance S. Two conceptions of machine phenomenality. Journal of Consciousness Studies 2007;14(7):154–66.
[6] Chrisley R. Philosophical foundations of artificial consciousness. Artificial Intelligence in Medicine, doi:10.1016/j.artmed.2008.07.011, this issue.
[7] Kuhn TS. The Structure of scientific revolutions. Chicago: University of Chicago Press; 1962.
[8] Lenggenhager B, Tadi T, Metzinger T, Blanke O. Video ergo sum: manipulating bodily self-consciousness. Science 2007;317(5841):1096–9.
[9] Aleksander I. The world in my mind, my mind in the world. Exeter: Imprint Academic; 2005.
[10] Aleksander I, Dunmall B. Axioms and tests for the presence of minimal consciousness in agents. Journal of Consciousness Studies 2003;10(4/5):7–18.
[11] Aleksander I, Morton H. Depictive architectures for synthetic phenomenology. In: Chella A, Manzotti R, editors. Artificial consciousness. Exeter: Imprint Academic; 2007. p. 67–81.
[12] Haikonen POA. You only live twice: imagination in conscious machines. In: Chrisley R, Clowes RW, Torrance S, editors. Proceedings of the symposium on next generation approaches to machine consciousness: imagination development intersubjectivity and embodiment (AISB'05 convention: social intelligence and interaction in animals, robots and agents). 2005. p. 19–25.
[13] Hesslow G. Conscious thought as simulation of behaviour and perception. Trends in Cognitive Sciences 2002;6(6):242–7.
[14] Hesslow G, Jirenhed D-A. The inner world of a simple robot. Journal of Consciousness Studies 2007;14(7):85–96.
[15] Ikegami T. Simulating active perception and mental imagery with embodied chaotic itinerancy. Journal of Consciousness Studies 2007;14(7):111–25.
[16] Chalmers DJ. Consciousness and its place in nature. In: Stich S, Warfield T, editors. The blackwell guide to the philosophy of mind. Malden: Blackwell Publishing Ltd.; 2003. p. 102–42.
[17] Gamez D. Progress in machine consciousness. Consciousness and Cognition 2007. doi: 10.1016/j.concog.2007.04.005.
[18] Metzinger T. Being no one: the self-model theory of subjectivity. Cambridge, MA: MIT Press; 2004.
[19] Seth AK, Baars BJ, Edelman DB. Criteria for consciousness in humans and other mammals. Consciousness and Cognition 2005;14(1):119–39.
[20] Heavey CL, Hurlburt RT. The phenomena of inner experience. Consciousness and Cognition 2008. doi: 10.1016/j.concog.2007.12.006.
[21] Metzinger T. Précis: being no one. PSYCHE 2005;11(5):1–35.
[22] Gallagher S, Zahavi D. The phenomenological mind: an introduction to philosophy of mind and cognitive science. London: Routledge; 2008.
[23] Baars BJ. A cognitive theory of consciousness. Cambridge: Cambridge University Press; 1988.
[24] Tononi G, Edelman GM. Consciousness and complexity. Science 1998;282(5395):1846–51.
[25] James W. The principles of psychology. New York: Henry Holt and Company; 1890.
[26] Franklin S, Graesser A. A software agent model of consciousness. Consciousness and Cognition 1999;8(3):285–301.
[27] Baars BJ, Franklin S. How conscious experience and working memory interact. Trends in Cognitive Sciences 2003;7(4):166–72.
[28] Dehaene S, Sergent C, Changeux JP. A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proceedings of the National Academy of Sciences of the United States of America 2003;100(14):8520–5.
[29] Wallace R. Consciousness: a mathematical treatment of the global neuronal workspace model. New York: Springer; 2005.
[30] Shanahan M. A cognitive architecture that combines internal simulation with a global workspace. Consciousness and Cognition 2006;15(2):433–49.
[31] Lau HC, Passingham RE. Relative blindsight in normal observers and the neural correlate of visual consciousness. Proceedings of the National Academy of Sciences of the United States of America 2006;103(49):18763–8.
[32] Block N. On a confusion about the function of consciousness. Behavioral and Brain Sciences 1995;18(2):227–87.
[33] Franklin S. A conscious artifact? Journal of Consciousness Studies 2003;10(4):47–66.
[34] Edelman GM. The remembered present. New York: Basic Books; 1989.
[35] Husserl E. On the phenomenology of the consciousness of internal time, translated from the German by JB Brough. Dordrecht: Kluwer Academic Publishers; 1991.
[36] Lakatos I. Falsification and the methodology of scientific research programmes. In: Lakatos I, Musgrave A, editors. Criticism and the growth of knowledge. Cambridge: Cambridge University Press; 1970.
[37] Shanahan M. Global access, embodiment and the conscious subject. Journal of Consciousness Studies 2005;12(12):46–66.
[38] Edelman GM, Tononi G. A universe of consciousness: how matter becomes imagination. New York: Basic Books; 2000.
[39] Seth AK, Edelman GM. Consciousness and complexity. In: Meyer B, Editor. Springer Encyclopedia of Complexity and Systems Science; in press.
[40] Edelman GM. Neural Darwinism: the theory of neuronal group selection. New York: Basic Books; 1987.
[41] Edelman GM. Naturalizing consciousness: a theoretical framework. Proceedings of the National Academy of Sciences of the United States of America 2003;100(9):5520–4.
[42] Seth AK, Baars BJ. Neural Darwinism and consciousness. Consciousness and Cognition 2005;14(1):140–68.
[43] Tononi G, Sporns O, Edelman GM. A measure for brain complexity: relating functional segregation and integration in the nervous system. Proceedings of the National Academy of Sciences of the United States of America 1994;91(11):5033–7.
[44] Sporns O, Tononi G, Edelman GM. Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. Cerebral Cortex 2000;10(2):127–41.
[45] Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 1969;37(3):424–38.
[46] Seth AK. Granger causality. Scholarpedia 2007;2(7):1667 [revision #37090].
[47] Tononi G. An information integration theory of consciousness. BMC Neuroscience 2004;5:42.
[48] Tononi G, Sporns O. Measuring information integration. BMC Neuroscience 2003;4:31.
[49] Seth AK, Izhikevich E, Reeke GN, Edelman GM. Theories and measures of consciousness: an extended framework.

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009

ceedings of the National Academy of Sciences of the United States of America 2006;103(28):10799—804.
[50] Bressler SL, Kelso JAS. Cortical coordination dynamics and cognition. Trends in Cognitive Sciences 2001;5(1): 26—36.
[51] Izhikevich EM, Edelman GM. Large-scale model of mammalian thalamocortical systems. Proceedings of the National Academy of Sciences of the United States of America 2008;105(9):3593—8.
[52] Nagel T. What is it like to be a bat? Philosophical Review 1974;83:435—50.
[53] Smith HO, HutchisonII CA, Pfannkoch C, Venter JC. Generating a synthetic genome by whole genome assembly: X174 bacteriophage from synthetic oligonucleotides. Proceedings of the National Academy of Science of the United States of America 2003;100(26):15440—5.
[54] Braitenberg V. Vehicles: experiments in synthetic psychology. Cambridge, MA: MIT Press; 1984.
[55] Seth AK, Dienes Z, Cleeremans A, Overgaard M, Pessoa L. Measuring consciousness: Relating behavioural and neurophysiological approaches. Trends in Cognitive Sciences 2008;12:314—21.

Please cite this article in press as: Clowes RW, Seth AK. Axioms, properties and criteria: Roles for synthesis in the science of consciousness. Artif Intell Med (2008), doi:10.1016/j.artmed.2008.07.009