Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

The evolution and development of consciousness: the subject-object emergence hypothesis

John E. Stewart

Evolution, Complexity and Cognition Group, Center Leo Apostel, Vrije Universiteit Brussel, Brussels, Belgium

ARTICLE INFO

Keywords:
Consciousness
Evolution of consciousness
Sensorimotor coordination
Subject-object emergence
Transitions in consciousness

ABSTRACT

A strategy for investigating consciousness that has proven very productive has focused on comparing brain processes that are accompanied by consciousness with processes that are not. But comparatively little attention has been given to a related strategy that promises to be even more fertile. This strategy exploits the fact that as individuals develop, new classes of brain processes can transition from operating 'in the dark' to becoming conscious. It has been suggested that these transitions occur when a new class of brain processes becomes object to a new, emergent, higher-level subject. Similar transitions are likely to have occurred during evolution. An evolutionary/developmental research strategy sets out to identify the nature of the transitions in brain processes that shift them from operating in the dark to 'lighting up'. The paper begins the application of this strategy by extrapolating the sequence of transitions back towards its origin. The goal is to reconstruct a *minimally-complex*, subject-object subsystem that would be capable of giving rise to consciousness and providing adaptive benefits. By focusing on reconstructing a subsystem that is simple and understandable, this approach avoids the homunculus fallacy. The reconstruction suggests that the emergence of such a minimally-complex subsystem was driven by its capacity to coordinate body-environment interactions in real time e.g. hand-eye coordination. Conscious processing emerged initially because of its central role in organising real-time sensorimotor coordination. The paper goes on to identify and examine a number of subsequent major transitions in consciousness, including the emergence of capacities for conscious mental modelling. Each transition is driven by its potential to solve adaptive challenges that cannot be overcome at lower levels. The paper argues that mental modelling arose out of a pre-existing capacity to use simulations of motor actions to anticipate the consequences of the actions. As the capacity developed, elements of the simulations could be changed, and the consequences of these changes could be 'thought through' consciously. This enabled alternative motor responses to be evaluated. The paper goes on to predict significant new major transitions in consciousness.

1. Introduction

1.1. *A hard problem for analytic philosophy*

Neither the methods used by analytic philosophy or by science have yet solved what Chalmers (1996) termed *The Hard Problem* of consciousness. Neither approach has yet explained convincingly why subjective experiences exists i.e. why it feels like something to be conscious; and why our conscious processes tend to 'light up' for us, rather than just operate 'in the dark' (like information processing in computers and in self-driving cars).

The Hard Problem also seems to have stood in the way of addressing fundamental issues about the evolution of consciousness: Why has consciousness emerged during evolution? What (if any) adaptive benefits are produced by conscious processing? Why does the realization of these benefits require consciousness? Answering these evolutionary questions would seem to require progress in overcoming The Hard Problem and the development of a functional understanding of consciousness and its effects.

In his 1996 book, Chalmers demonstrated that The Hard Problem is indeed a very hard problem for analytic philosophy. The considerable efforts made by analytic philosophers to solve the problem since then have confirmed how difficult it is for those equipped only with the methods of analytic philosophy. In 2022, consciousness is still a Hard Problem for analytic philosophy (e.g. see White, 2021). This lack of progress by analytic philosophy is not surprising. Progress in human understanding of natural phenomena has been driven primarily by hypothesis-guided experimentation and the related methods of science.

E-mail address: future.evolution@gmail.com.

https://doi.org/10.1016/j.biosystems.2022.104687
Received 17 January 2022; Received in revised form 24 April 2022; Accepted 25 April 2022
Available online 30 April 2022

In contrast to science, analytic philosophy places much greater emphasis on critical analysis. As such, analytic philosophy has played only a very minor role in the progress of the empirical sciences in the last century (e.g. see Unger 2014).

1.2. Hypothesis-guided experimentation – comparative methods

However, as I will substantiate in detail in this paper, the problem of consciousness is not such a hard problem for the methods and tools of science. The development of an understanding of consciousness is a challenge for science, but it is no harder than a number of other complex challenges that science has met successfully in the past. Scientific inquiry has demonstrated repeatedly that problems which cannot be resolved by analytic philosophy alone are amenable to methods that rely on hypothesis generation and empirical testing (I discuss this issue in greater depth from a philosophical perspective in Section 8.2 below).

In particular, science has a strong track record of identifying the functional significance of subsystems of organisms, understanding how these subsystems are organised to serve these functions, and explaining how they produce their various adaptive effects (including any emergent effects). A research strategy that has proven particularly powerful for these purposes is to compare and contrast a subsystem of interest with similar subsystems that are less complex and better understood. Ideally, the subsystem of interest will differ from the simpler subsystem only in relation to a small number of additional features that researchers are attempting to explain. Experimentation can then focus on identifying the additional functionality provided by these features and on how this extra functionality is organised and produced.

Such a research strategy can be even more effective when the less-complex, 'better-understood' subsystem is an evolutionary or developmental precursor of the subsystem of interest. As I will argue in detail below, the science of consciousness has not yet fully exploited the potential of such an evolutionary/developmental research strategy.

However, a research strategy that uses a limited version of the comparative approach has already proven very productive in investigating the consciousness subsystem in organisms, particularly humans. This approach takes advantage of the existence in the human brain of adaptive processes that operate outside of conscious awareness (i.e. in the dark). This research strategy involves comparing and contrasting these processes with ones that operate with some degree of conscious awareness. In these circumstances, hypothesis-guided experimentation has been able to focus on what distinguishes conscious from unconscious adaptive processes and the functional significance of these differences.

1.3. The success of existing comparative methods – Global Workspace Theory

Such a comparative approach has been responsible for much of the progress made by the research program into the functioning of consciousness that has been the most successful to date. This program encompasses Global Workspace Theory (Baars, 1988; Baars et al., 2013) and the related Global Neuronal Workspace Theory (Dehaene et al., 1998; Dehaene and Naccache 2001; Mashour et al., 2020). I will refer to these collectively as the GWT research program.

The GWT research program relies heavily on what Baars refers to as *contrastive analysis*. This is a more specific form of what I term the comparative approach. Using this research strategy, the GWT program has made considerable progress towards identifying the additional functionality associated with conscious processes and how this functionality is produced. The research program has shown that in some circumstances, this additional functionality can confer adaptive capacities that are superior to those provided by adaptive processes that proceed unconsciously. Broadly, GWT suggests that conscious processes involve the use of *global broadcasts* about adaptive challenges to recruit specialist *agents* from across the brain. Appropriate combinations of these agents can contribute their specialist skills to devise novel adaptive responses to the challenges that have been broadcast. This capacity is particularly useful for developing responses to adaptive challenges that have not been encountered before and are not the subject of pre-existing learned or innate responses i.e. where there are uncertainties and ambiguities about how to adapt.

But despite its successes, GWT has not yet produced a widely-accepted explanation of why a global broadcasting and recruiting system would necessarily generate conscious experience. It has not developed a plausible hypothesis that can answer the fundamental question: why would the information processing embodied in the global workspace system somehow feel like something, rather than just proceed in the dark? Why would the contents of the global broadcasts light up?

This paper sets out to overcome this and other limitations of GWT and other functionalist theories that attempt to explain consciousness. It does so by expanding and extending the 'comparative' research strategy that has proven powerful in elucidating the functioning of subsystems in organisms in general, and that has produced most of the progress that has been made in understanding consciousness to date.

2. Designing an evolutionary/developmental research program

2.1. Developmental transitions in consciousness

The potential to expand the comparative research strategy is considerable. This is because relatively little attention has been given by consciousness research to developmental approaches. A developmental approach can be particularly productive because as humans develop, psychological processes that previously operated in the dark can become conscious (Kegan 1982, 1994). These developmental transitions provide numerous opportunities to explore the nature of the changes in functioning that produce the shifts to conscious processing.

Kegan identifies the key changes that occur when a developing individual undergoes such a transition: processes that were part of the subject at an earlier stage in development becomes object to a new, higher-level subject at the next developmental stage. In other words, these transitions in human development occur when what previously operated unconsciously in the dark, now lights up for an emerging, higher-level subject, and feels like something to have.

Kegan notes that processes that operate as part of the subject proceed automatically and are not able to be influenced consciously. But this changes significantly when a class of psychological processes such as emotions or a particular level of thinking moves from subject to object—the processes can then be subject to what is experienced as 'conscious choice'. When psychological processes become object to consciousness, the individual is no longer bound by them automatically. Instead, individuals experience themselves as having a degree of psychological distance from them. They experience themselves to some extent as 'standing outside' the processes.

Drawing on the terms used by a number of the spiritual and contemplative traditions, an individual tends to be 'non-attached' to processes that have moved from subject to object, or 'dis-identified' from them (e.g. see Stewart 2007). As individuals develop psychologically, the classes of psychological processes that they experience as object expands, and they tend to experience themselves as having greater psychological freedom (e.g. see Stewart 2017).

A specific example given by Kegan is when emotions move from being part of the subject to become object to a newly-emerged, higher-level subject. He observes that when an individual undergoes this developmental step, the individual moves from being a person whose emotions *have them*, to one *who has* emotions (and who therefore can choose whether or not to 'go with' particular emotions as they arise). Such an individual becomes non-attached to their emotions, gains some psychological distance from them, and is no longer bound by them.

2.2. Transitions in consciousness during meditation and evolution

These are not the only circumstances where individuals may experience a shift in which psychological processes that previously operated largely in the dark become conscious. These kinds of shifts also occur during the practice of meditation (for overviews see Stewart 2007; Combs 2009). In meditation, it is commonplace for individuals to move back and forth between states in which they are bound up in thoughts and emotions, and states in which they experience thoughts and emotions as objects arising in consciousness. The meditator moves between states in which thoughts and emotions *have them*, to states in which *they have* thoughts and emotions. In fact, meditation can be intentionally used as a practice directed at developing the capacity to move psychological processes from subject to object, from processes that occur in the dark to ones that are object to consciousness.

It is also reasonable to adopt the working hypothesis that similar transitions from non-conscious to conscious processing have also occurred as biological organisms have evolved.

2.3. A comparative research strategy focusing on transitions in consciousness

The existence of these shifts during evolution, development, and meditative practices opens up broad new research possibilities. They provide potentially fertile opportunities to explore differences in functioning between conscious and non-conscious processes, and to understand how these differences are produced. Key new research questions that are highlighted by these phenomena include: What are the changes in the organisation of systems within the brain that result in particular psychological processes moving from subject to object? In what ways do the transitions from subject to object provide additional adaptive functionality for the organism? What is it about the ways in which this additional functionality is produced that results in these psychological processes lighting up and feeling like something to have?

2.4. Beginning with a minimally-complex subject-object subsystem

Kegan has not developed a theory or undertaken experimentation that addresses these key questions in detail. However, his suggestion that cognitive and social/emotional development is characterized by a sequence of developmental stages suggests a simple starting point for this research program. Kegan notes that before the developmental trajectory begins, no psychological processes are conscious (all is subject, nothing is object, as he puts it). As development proceeds, more and more processes become object to consciousness, until eventually all those processes that are capable of being conscious are object. A potentially productive starting point for the research program would be to attempt to model and investigate the first step in this developmental sequence. This step would encompass the first emergence of psychological processes that are object to a subject. Such an approach has the potential to facilitate the reconstruction of a minimally-complex (and maximally tractable) model of the first emergence of conscious processing. Such a model would have the potential to explain the emergence of conscious processes from processes that were previously unconscious.

2.5. Avoidance of the homunculus fallacy

I emphasize here that this focus on such a minimally-complex subsystem also has the potential to ensure that the research program does not commit the *homunculus fallacy* (Dennett 1991). This fallacy arises when an attempt to account for consciousness relies on the existence of a homunculus within the brain that is assumed to be fully conscious and capable of perceiving representations of the external environment that are presented to it. This assumption is made without any further explanation about how this homunculus/subject is able to exhibit the consciousness that the hypothesis is attempting to explain. It is treated largely as a 'black box'. By itself, such a model is obviously incapable of explaining conscious functioning. Furthermore, any attempt to overcome this difficulty by hypothesizing that the emergent subject is capable of conscious processing because it contains a subject that is conscious, and so on, produces an infinite regress. However, commencing the evolutionary/developmental research strategy by focusing on a 'minimally-complex' model of the emerging subject has the potential to avoid this fallacy. This is because the functions that are embodied in such an emerging subject are likely to be far more simple and understandable than those embodied in a subject that emerges at higher levels of development. If these simpler processes can be sufficiently understood and explicated in their own right without reliance on conscious black boxes, the fallacy has been avoided. The minimally-complex model that I reconstruct below exploits this potential successfully.

2.6. The learning capacities of pre-conscious organisms

Adopting Kegan's starting point, we need to commence our model building with a hypothetical organism that adapts only through processes that are non-conscious. The organism has no conscious experience whatsoever– its adaptation occurs entirely in the dark. This immediately raises the question: what are the processes that enable organisms to discover novel adaptations without the involvement of consciousness? Once we have identified these non-conscious processes, we can assess their limitations, and explore how the limitations might be overcome by the emergence of a minimally-complex subject-object subsystem.

It is not straightforward to clearly distinguish adaptive processes that are conscious from those that are non-conscious by considering only human experience. The highly-developed and complex forms of consciousness in humans appear to be involved at least to some extent in most human adaptive processes (Baars 1988). It is not easy to disentangle conscious from unconscious processes in humans, except in some clear-cut cases (e.g. the adaptive processes associated with homeostasis appear to operate mostly outside consciousness (Solms, 2021). Furthermore, the alternative of focusing on the experience of non-human animals is no easier—they are unable to report whether their adaptive processes operate in the dark or not.

However, there is a way forward. The difficulty can be resolved at least provisionally by considering artificial intelligence that is capable of complex learning, but is generally accepted as lacking consciousness. A particularly clear example is the artificial intelligence embedded in self-driving cars and in robotics. This AI is clearly capable of using complex learning processes to discover novel adaptations (e.g. see Pierson and Gashler 2017). These learning processes are analogous to the simpler forms of learning that operate in living organisms. Yet it is relatively uncontroversial that these learning processes operate in the dark in self-driving cars and robotics. This can reasonably be assumed to hold true even if the learning processes were actually instantiated rather than digitally simulated. On this basis, we will continue the development of our research strategy on the assumption that comparable learning processes in biological organisms should equally be capable of operating in the dark.

The analogous learning processes embodied in biological organisms can be considered in two categories: reinforcement learning and associative learning. With reinforcement learning, the organism learns to perform actions that are positively reinforced or rewarded (and learns not to perform actions that are negatively reinforced). Before an organism has learned or inherits an action that is appropriate in particular circumstances, it will tend to search for an effective response by trial-and-error. Skinner (1981) described this form of learning as 'selection by consequences'.

The type of associative learning that has been most studied in organisms is classical/Pavlovian conditioning. But associative learning also includes any other process in which the organism learns that separate stimuli are associated and related (Turchin 1977). Through associative learning, an organism is able to discover correlations, regularities and patterns in space and time in its environment, including part/whole relationships.

As I have indicated, it seems generally accepted that these two categories of learning processes operate without conscious experience in self-driving cars and in other forms of intelligent AI. And I have suggested that it is equally reasonable to assume that these learning processes are capable of operating without conscious experience, at least in organisms that lack highly-developed and complex capacities for consciousness. I will proceed here on the further reasonable assumption that when these learning processes first emerge in living organisms, they operate without consciousness. This 'working assumption' is made in relation to emergence during development as well as for evolutionary emergence.

3. The emergence of a minimally-complex subject-object subsystem

3.1. Application of the evolutionary/developmental research strategy

We are now ready to begin our reconstruction of the emergence of a minimally-complex subsystem that is capable of conscious experience. We commence the reconstruction with an organism (or AI) that is capable of complex learning (both reinforcement/operant learning and associative learning), but without conscious experience. We will then adopt an evolutionary/developmental perspective to generate and investigate the following key questions:

(1) In what ways would the adaptive capabilities of such a non-conscious organism be limited?
(2) How could an organism overcome at least some of these limitations by the emergence of a conscious subsystem i.e. by a subsystem in which some processes become object to an emergent subject?
(3) Using a reverse engineering approach, how would such a conscious subsystem need to be instantiated and organised if it were to fulfil this potential to enhance adaptive capabilities? What kind of conscious, minimally-complex subsystem could take advantage of these adaptive affordances?

3.2. The adaptive limitations of non-conscious organisms

First, we will identify the adaptive limitations that would beset such a non-conscious organism (or AI). We begin by noting that a fundamental limitation of reinforcement learning (including operant conditioning) is that it is costly. It discovers adaptive behaviours by processes that involve a degree of trial-and-error—the organism tries out different behaviours until one is found that achieves the organism's adaptive goals and is therefore rewarded. Although this is a costly process, it is worth noting here that it is far less inefficient and costly than the discovery of adaptations through gene-based natural selection. Natural selection operates across the generations through the differential survival of individuals. In contrast, operant-like learning processes operate far more quickly during the lives of individuals (Dennett 1996).

In these circumstances, we can reasonably expect that the emergence of a capacity to discover adaptations that further reduces or eliminates the need for costly trial-and-error would be strongly favoured by evolution (and by learning processes). This leads to the following questions: Is there something approaching a form of 'one-shot' learning or adaptation in real-time that could have emerged and overcome the limitations of reinforcement learning, at least in some circumstances? Would this capacity require the emergence of a subject-object subsystem?

3.3. Overcoming these limitations through real-time sensorimotor coordination

The kind of emergent subsystem that we are attempting to reconstruct would appear to have to meet the following requirements:

(1) The emergent subsystem would be able to be realized simply. It would begin as a very minor subsystem within the organism's pre-existing adaptive systems which all operate in the dark. The emergent subsystem would be likely to have simple origins because complex new adaptive systems cannot suddenly leap into existence, fully formed and operational. For example, modern humans are capable of achieving one-shot learning through the use of mental models. These mental simulations can be used to predict what kinds of actions will be adaptive in circumstances that may not have been encountered before. But such arrangements are far too complex to plausibly leap into existence in the circumstances under consideration here.

    This requirement for simplicity is also consistent with a research strategy that focuses on the reconstruction of a conscious subsystem that is minimally-complex.

    The sequence of developmental stages in humans that has been identified by Piaget and others suggest that the most likely candidate for the emergence of a minimally-complex subsystem would involve processes at the sensorimotor level (e.g. see Piaget 1969);
(2) The emergence of the subsystem would build incrementally upon processes that already exist in non-conscious organisms, and would be able to be produced by adaptive mechanisms that are within the capabilities of such organisms;
(3) It is likely that the new adaptive capacity that is produced by the conscious subsystem would also be associated with consciousness in humans;
(4) Consistent with Kegan's model, the architecture and functioning of the emergent subsystem would comprise a subject-object form of organisation; and
(5) Ultimately, the emergent subsystem and its functioning must at least be consistent with the hypothesis that some of its processing feels like something to the subsystem.

A prime candidate that seems capable of meeting these requirements is what I will refer to as *Real-time Sensorimotor Coordination*. This is a coordination processes in which feedback from sensory representations is used in real time to guide motor actions to produce adaptive outcomes in the organism's environment. A familiar example is hand-eye coordination. Importantly, hand-eye coordination in humans significantly reduces the extent to which we need to use costly trial-and-error operant processes in order to discover how to achieve particular outcomes in the physical world. Take as an example the movement of an object to a specific location in relation to other objects. The actual sequence of motor actions that are needed to achieve this goal does not have to be discovered by trying out various movements until a sequence that works is eventually hit upon. Instead, the task is achieved by the perceptual monitoring in real time of the effects of 'voluntary' motor actions on the achievement of the task. This perceptual monitoring is used to guide the actions in real time so that they produce the desired outcome.

Imagine undertaking an intricate hand-eye coordination task such as threading a needle. When we do so, we find that attempting the task requires us to give very close and continuous attention to our fingers, the needle and the thread. This enables us to become continuously aware of detailed analogical representations of our attempts to insert the thread through the eye of the needle. These representations in turn provide us with real-time feedback about the effectiveness of movements we make with our fingers. We use this feedback to guide us as we complete the task.

It is instructive to also imagine how our experience of threading a needle would change in the absence of a subject-object system i.e. without a subsystem that inspects relevant representations and generates real-time feedback to guide motor actions. We would still have vision and the representations that it generates. But there would be no subject to interpret and use the representations in real time to provide us with immediate feedback. In the absence of this real-time feedback about progress toward attaining our goal, we would have to rely only on the actual achievement of the final goal to inform us when we have accomplished the task. We would receive feedback only in the form of the operant reinforcement that accrues only when the task is fully completed. As a consequence, in the absence of prior learning that is specific to the particular circumstances, we would have to rely on trial-and-error to discover what to do to thread the needle.

We have all experienced what it feels like to perform hand-eye coordination tasks in the absence of prior learning. We know that our hands and the objects we manipulate light up in our awareness when we do so. It is difficult to imagine that these coordination tasks could be performed effectively in the absence of prior learning without such real-time awareness of our hands and the objects we are manipulating.

3.4. The reconstruction of a subject-object subsystem capable of real-time sensorimotor coordination

These intuitions and subjective experiences about hand-eye coordination are suggestive that conscious subject-object subsystems might first have arisen due to their capacity to enable Real-Time Sensorimotor Coordination. However, these observations are incapable by themselves of constituting an adequate explanation of the emergence of consciousness. So I will now turn to substantiating in detail how real-time sensorimotor coordination could be enabled by the emergence of a minimally-complex subject-object subsystem. Furthermore, I will demonstrate that the organisation and functioning of this subsystem supports the hypothesis that particular components of the subsystem have conscious experience.

Consistent with the research strategy I have developed to this point, I will begin by considering how a minimally-complex subsystem would need to be constituted if it were to be capable of Real-Time Sensorimotor Coordination. The goal of the strategy is to attempt to re-construct something like the first subject-object subsystem that emerged in the evolutionary transition from pre-conscious to conscious organisms. Such a subsystem is also likely to be similar to the first conscious subsystem that emerges in organisms that undergo a comparable developmental transition during their lives.

So we begin with a pre-conscious organism that has the following characteristics: it is capable of processing sensory inputs; this processing can produce internal representation of aspects of its environment; these representations may be more-or-less analogical (for example, the spatial relations between components of the representations may reflect corresponding spatial relations in the external environment); these analogical representations may be produced by sensory processing that maintains the relations within the retinal image, and/or by neural networks that secondarily reconstruct the image in an analogical form (Shen et al., 2019); and the organism is enabled by operant learning and associative learning to discover motor actions that are adaptive given particular sensory inputs and representations.

These learning capacities would enable the pre-conscious organism to learn how to adaptively control the orientation of its attention and gaze. This is an important capacity: the ability to direct attention can contribute significantly to the adaptability of an organism. This is because images that fall on the retina's fovea are processed in much greater detail than other components of the image. So there is adaptive benefit in orienting the eye so that the images that are most relevant to the organism are subjected to this deeper processing, and are tracked as they move and/or as the organism moves (Graziano and Webb 2016). Optimal orientation could be achieved by motor actions that, for example, change the position of the body, head and/or eyes.

Such a pre-conscious organism would be capable of learning to identify the kinds of stimuli that can be given attention beneficially in particular circumstances. It could also learn by trial and error the particular motor actions needed to direct gaze and attention at relevant stimuli, including the actions needed to track movement of the stimuli. These learned motor patterns could include heuristics and implicit models that are able to generate a class of adaptive actions However, learning would not be necessary where the organism is already 'hard-wired' with innate responses as a result of the operation of evolutionary processes (see Berthenthal, 2020). All of this processing could occur in the dark.

There would be considerable adaptive benefit in the emergence of a subject-object subsystem that is capable of moving attention to the most salient features of the environment and tracking them, without depending on wasteful trial-and-error learning (I use *salient* here broadly to include features and stimuli that are goal-relevant). In real time, such a subsystem could coordinate the particular motor actions needed to direct attention at particular features and to follow them as necessary i. e. to 'grasp' and 'examine' them. The organism would no longer have to learn the required motor actions predominantly by trial and error. We have seen that hand-eye coordination enables objects to be manipulated in real time in circumstances for which the organism has not already learned (or inherited) appropriate motor-action programs. On a similar basis, the real-time coordination of attention would enable attention to be optimally disposed in circumstances for which pre-existing adaptive responses are not available.

These considerations suggest that a prime candidate for a minimally complex subject-object subsystem is a subsystem that coordinates in real time the disposition of attention towards relevant environmental stimuli. I will now explore this possibility in greater detail.

3.5. The architecture and functioning of a subject-object subsystem capable of the real-time coordination of attention

What are the key functional features that would need to be exhibited by a subject-object subsystem that coordinates attention with relevant stimuli in real-time? A reverse-engineering approach suggests that the subsystem would need to include the following functional processes:

(1) **In order for the subsystem to be able to take over the control of attention when it is adaptive to do so, the subsystem would need to be able to inhibit motor actions evoked by current sensory representations.** This would enable the subsystem to substitute alternative motor actions, and to implement sensorimotor coordination of attention in real time.

The capacity of the subsystem to inhibit other responses would enable current representations to become object to the emerging subsystem, in the sense developed by Kegan—the representations would no longer automatically evoke learned or innate motor actions that unfold in the dark. Instead, the subsystem would 'stand outside' these processes and would not be bound by them – it could 'choose' alternative actions. For example, the capacity to suppress automatic responses would enable the subsystem to give uninterrupted attention to representations for as long as is necessary to coordinate attention with relevant stimuli in real time.

This inhibition of automatic motor actions is likely to be adaptive in circumstances where there is uncertainty about how the organism might act i.e. where there are no pre-existing learned or innate responses that are likely to be adaptive, given the previously-accumulated adaptive repertoire of the organism. This inhibition would also be able to suspend operant searching.

(2) **In order to be able to use representations in real time to guide attention to where it is most adaptive, the subsystem would need to be able to make use of analogical and other information that is embodied in the representations.**

However, the subsystem would need to extract only limited amounts of information from representations in order to guide attention toward relevant stimuli in real time. It would not have to reprocess and re-interpret all the information produced by the organism's sensory processing. It would not have to embody a homunculus, or anything like one. Instead, the subsystem would need only a more-or-less analogical sketch of the outputs from the visual processing system. The sketch would not need to be any more detailed than is necessary to enable the subsystem to monitor the disposition of attention in the environment and track attention as it is moved by the subsystem.

The minimally-complex subject-object subsystem would not perform the function of assessing and utilizing all the sensory information that might be adaptively useful to the organism. These functions would continue to be performed by processes that operate in the dark, as they are in non-conscious organisms. The fact that conscious subsystems do not process detailed representations as they guide attention is consistent with what is known as *The Grand Illusion*. Humans report that they experience in rich detail the contents of their visual field that are outside the small spotlight of attention. However, experimental evidence indicates that they are not in fact aware of this detail (Noë and O'Regan 2000). The limited detail that is actually available to consciousness is, however, sufficient to be used to scan the environment, identify salient features, and guide the movement of attention to them.

(3) **In order to be able to use visual representations to deploy attention adaptively in real time, the subsystem would also need to be able to identify the particular motor actions needed to coordinate attention with relevant stimuli.** This would require a coordination process that operates on a similar basis to hand-eye coordination. The subsystem would need to use real-time information about the impacts of its motor actions. It would use sensory representations of the impacts of its actions to obtain feedback in real time about the appropriateness of particular actions. This would enable the subsystem to identify and initiate the particular motor actions that are shown by visual representations to be necessary to move attention to where it is most beneficial. As discussed, these motor actions could, for example, include movements of the whole body of the organism and/or movements of the head or eyes.

Significantly, from the perspective of the emerging subject-object subsystem, these motor actions are initiated within the subsystem itself, and are controlled and chosen by it.

(4) **The processes within the subsystem that stand outside, observe and use the representations to guide attention constitute the subject.** The representations would be object to this subject, in the sense discussed by Kegan.

The Subject-Object Emergence Theory hypothesizes that the subject will experience itself as being aware of the representations and aware of initiating and controlling the motor actions that dispose attention. Furthermore, the subject will experience the motor actions as voluntary. As the subject attempts to adjust its motor actions to achieve particular outcomes (utilizing feedback from its visual representations), it will experience itself as choosing particular motor actions over other motor actions that appear equally possible. It will experience a simple form of apparent agency and free will.

When we watch a human baby lying on its back trying to grasp and manipulate a rattle dangled above it, it appears we are observing the actual development of a capacity for hand-eye coordination. We can intuit that we are watching the emergence of a conscious subsystem that embodies an experience of voluntary control. Furthermore, if we have the capacity to observe our own usage of conscious hand-eye coordination in real time, we can directly experience the operation of such a subsystem and the role played within it by its various component processes. For a detailed treatment of the emergence of voluntary action, see Gunji et al. (2017).

(5) **In order to identify where attention should be directed to achieve the greatest adaptive benefit, the subject-object subsystem would need to receive input about the salience of particular features in the organism's environment.** The pre-conscious organism is already equipped with processes that assess salience and determine priorities amongst competing saliences. These processes operate in the dark and would continue to do so in an organism which includes only a minimally-complex subject-object subsystem. Included in these processes are ones that are more-or-less hard wired by evolutionary processes, as well as processes that are learned by associative learning (including by classical conditioning). As a subject-object subsystem emerges, these processes perform the role of informing the subsystem of those features of the environment that are adaptively significant and warrant attention. The processes will also inform the subsystem when a particular object of attention has been sufficiently investigated and when attention could usefully be moved elsewhere.

As we have noted, when attention is given to a salient object of perception (uninterrupted by learned and innate responses), more-detailed visual information and processing about the object becomes available to the organism. If there is uncertainty and ambiguity about what adaptive response would be effective in the circumstances, and if the subsystem inhibits other responses, deeper visual processing may continue (supplemented by shifts of attention to other perceptions that may reduce the uncertainty). This deeper visual processing would also include further processing within the neural networks of the organism (Graziano and Webb 2016). In particular, this deeper processing would be able to draw on networks of associations that have been learned in the past. Exploration of these networks would have the potential to identify any associations that might be relevant to the current adaptive challenge.

This exploration could occur through the process known as *spreading activation* (e.g. see Heylighen and Bollen 1996). Activation that spreads across a neural network will tend to preferentially follow the strongest linkages that have been established by previous associative learning (associative learning tends to strengthen links that are frequently activated). In this way, spreading activation has the capacity to draw on past learning to discover additional associations that might be relevant.

However, it is worth noting here that the wider activation that might be produced when prolonged attention is given to a particular percept is not the kind of global broadcasting process envisaged by some models of GWT (Baars 1988). Spreading activation is not an undirected process that communicates with all other brain processes equally. It does not involve indiscriminate broadcasting. Instead, as activation spreads throughout the brain, it gives priority to exploring those linkages in the brain that have been strengthened by past learning.

Spreading activation may discover particular associations that resolve the uncertainty and ambiguity facing the organism. For example, it may evoke pre-existing learned responses that are adaptive in the circumstances. When such an association is found that resolves the adaptive uncertainty, and when a previously-learned adaptive response is evoked, attention can be moved to some other salient percept (consciously or unconsciously, depending on the circumstances). Once the uncertainty is removed, the organism will 'know what to do', based on its previously-learned associations and reinforcement learning. When this occurs, the organism's salience landscape will change, and this will be fed back to the subject-object subsystem.

Against this background, it is worth noting that the functioning of the subject-object subsystem would not occur within a physically identifiable module that is located in any particular place in the brain. Like a number of the processes that it co-opts, it would be a distributed subsystem.

Equipped with these functions, the subject-object subsystem would be able to use the organism's internal sensory representations to scan the environment and to move attention directly to salient phenomena and to track them, all in real time. The organism would not have to rely primarily on trial-and-error processes to discover the actions needed to locate salient features in the environment and to shift attention to them. A subject-object subsystem that has these capabilities will be adaptively superior in circumstances where the organism does not already possess learned or innate motor actions that will adapt the organism effectively. An organism equipped with such a subject-object subsystem will be capable of the real-time sensorimotor coordination of attention.

As noted briefly, the ability of the subject-object subsystem to control attention (and therefore perception) also opens up further new adaptive possibilities. The subsystem is able to direct attention at a particular object, uninterrupted by other behaviours. This enables deeper internal processing that uses spreading activation to search previous associations and learnings to discover appropriate adaptive responses. Spreading activation can explore existing associative networks to recruit resources that are relevant to dealing adaptively with uncertainty facing the organism.

In these ways, the subject-object subsystem controls both *external attention* (the disposition of the eyes and fovea relative to the environment), as well as *internal attention* (the engagement of deeper processing across the brain in the search for previous experience that may be relevant to current adaptive challenges).

As we have seen, the superior capacities of such a subject-object subsystem will enable the organism to orientate attention effectively in novel circumstances without having to rely primarily on trial and error. But once the organism discovers how to orientate attention adaptively in a particular set of circumstances, this will become a learned behaviour that can be deployed again in those circumstances without conscious involvement. Increasingly, the organism will accumulate learned responses that are adaptive in specific situations. These learned responses will then tend to be evoked automatically when the relevant circumstances are encountered again in the future. The acquisition of learned responses will tend to obviate the need to engage the subsystem again in the circumstances where the learned responses are relevant. As a consequence, if we observe how our own attentional system functions, we find that it mostly operates unconsciously, using the automatic, learned responses that have been discovered and accumulated with the aid of conscious processes in the past.

The adaptive superiority of the subject-object subsystem in circumstances where pre-existing learned responses are not available means that selection will tend to favour its emergence, and motor actions that contribute to its functionality will tend to be reinforced positively.

3.6. What would it be like to be such a subject-object subsystem? - Consciousness and voluntary control

The Subject-Object Emergence Theory hypothesizes that the emerging subject will experience as objects of awareness the sensory representations that it scans and utilizes. Furthermore, the subject will experience as voluntary its use of these representations to guide its actions. It will experience itself as making voluntary choices about its motor actions. In contrast, the emerging subject will experience the operation of learned and innate motor actions and responses as non-voluntary and outside its immediate control.

An alternative, competing hypothesis is that the processes that constitute such a subject-object subsystem could proceed in the dark, without entailing any conscious experience. But the methods of science do not require that this alternative hypothesis be given a priori any superior status to the Subject-Object Emergence Hypothesis. A scientific approach does not require this alternative to be accepted as some kind of default position that has to be disproven analytically before the Subject-Object Hypothesis can be taken seriously. Instead, a 'Popperian' approach to scientific enquiry requires that the competition between these alternative hypotheses be resolved by subjecting their respective predictions to testing – i.e. by assessing their ability to make novel and bold predictions that can be falsified empirically (Popper 1959; Solms, 2021; Thornton 2021).

We will return to this issue of testing the Subject-Object Emergence Hypothesis in Section 8 below. We will now begin to explore how the Subject-Object Emergence Theory can be extended in order to account for the evolution and development of forms of conscious processing that are more complex than sensorimotor coordination.

4. The further evolution/development of consciousness: the emergence of a capacity for concrete modelling

4.1. Extending the research program

As discussed above, Kegan suggests that as humans develop, an increasing variety of brain processes become object to consciousness. This occurs through a recursive developmental process in which brain processes that are initially part of the subject (and operate outside conscious awareness) become object to a new subject which is at a higher level. It is likely that a similar recursive process that produces a hierarchy of levels has also driven the evolution of consciousness.

The evolutionary/developmental research strategy can be extended in order to reconstruct how consciousness might evolve and develop after its first emergence in a minimally-complex subject-object subsystem. Key goals of this extended strategy are to:

(1) Identify the adaptive limitations of the pre-existing level (these limitations are adaptive affordances when viewed from the perspective of the evolutionary and learning processes that adapt organisms);

(2) Reconstruct changes to the subject-object subsystem that would overcome these adaptive limitations to some degree, including by making additional processes object to a new, higher-level subject.

    Previously, these additional processes will have operated automatically, outside awareness. But when the processes become object, they can instead be adapted consciously by actions that are experienced as voluntary. Because the changes to the subsystem improve adaptability, they will tend to be favoured by evolutionary and learning processes.

(3) Develop a reconstruction in which changes to the subsystem are produced by incremental, minimally-complex changes to the existing subsystem and associated processes, not by the sudden emergence of complex novel processes.

It is beyond the scope of this article to apply this research strategy in detail to identify all the levels in the evolution/development of consciousness that follow the initial emergence of a simple subject-object subsystem. Instead, I will focus on four subsequent levels that are particularly significant in human development. Three of these levels concern cognitive development, and the fourth involves social/emotional development (for alternative attempts to identify levels in the psychological development of humans, see Piaget 1969; Fischer, 1980; Kegan 1982; Commons et al., 1998; Mascolo 2015).

I will refer to the first of these levels as the *Concrete Modelling Level*. Broadly, it deals with cognitive processes that underpin a number of the key capacities that characterize Piaget's *Concrete Operations Stage* of psychological development (Piaget 1969).

4.2. The limitations of previous levels

Consistent with the research strategy outlined above, I will begin my consideration of the emergence of Concrete Modelling by identifying key adaptive limitations of the pre-existing level (i.e. the level at which conscious processing is largely confined to sensorimotor coordination). As pointed out by Vygotsky (1978), organisms limited to this level are largely 'slaves to their visual field'. Their adaptation is largely restricted to taking into account only the circumstances that are represented in their sensory fields. Their conscious processing cannot adapt effectively to circumstances that are not being sensed currently, including future circumstances. Their conscious processing is unable to 'go offline' and simulate possibilities that are not currently represented in their sensory fields.

A particularly vivid example of this limitation can be seen in the behaviour of a dog which is apparently attempting to avoid punishment by its owner: the dog may put its head under a couch, preventing it from seeing its angry owner, but with the rest of its body fully visible. Because its conscious processing is unable to go off-line and create internal representations of its master standing behind it, the dog behaves as if it has successfully evaded the attention of its owner.

4.3. Overcoming the limitations – the emergence of a capacity for internal simulation/modelling

How could the further development of the subject-object subsystem overcome these limitations, at least in part? More specifically, how could this be accomplished by a shift in which brain processes that previously operated in the dark, become object to consciousness? How could this enable the subject-object subsystem to go offline relative to sensory processing, 'simulate' new adaptive possibilities and escape slavery to sensory fields?

In order to address these issues, it is first necessary to consider particular brain processes that have not yet been included in this discussion. I am referring here to processes that have been examined in some detail by the rapidly expanding research field associated with what is known as Predictive Processing. This field is a formalisation of the long-held understanding that sensory processing is not restricted to the use of immediate sensory inputs alone. Instead, the representations that are used by organisms also take into account the previous experiences and learnings of the organism. The integration of these disparate sources of information produce the organism's 'best hypothesis' about the state of salient aspects of its environment, given sensory inputs *and* past experience (e.g. see Hawkins 2004; Friston 2010).

Central to predictive processing are the processes that anticipate the sensory and other effects produced by motor actions. These processes use internal 'simulations' of motor actions to predict the sensory consequences that will result from the actions. If the simulated predictions are not met, the organism can modify its actions until the outcomes of its actions enable it to achieve its goals.

Drawing upon a predictive processing framework, Pezzulo (2011) argues persuasively that this ability to use internal simulations/models can be seen as an early step in the emergence of a more comprehensive capacity that would significantly enhance adaptability. Initially, the organism simulates only a single motor action and its consequences. But as this capacity develops, the organism could acquire an ability to simulate an array of alternative motor actions, including their predicted impacts on the organism and its environment. Such an extended capacity to produce simulations could be used to identify and select motor actions that are predicted to be adaptively superior. This ability to utilize an expanded range of simulations could also be used to model adaptive possibilities in circumstances that are not represented directly in the organism's sensory fields (including in possible future circumstances). The organism could use these simulations/models to discover beneficial adaptations that are not accessible to organisms without this capacity.

We can use the subject-object emergence framework to explore how such a simulation/modelling capacity could develop, and its consequences for the conscious experience of the organism.

4.4. Overcoming the limitations – new subject-object emergence

Subject-Object Emergence Theory suggests that the development of this capacity will require that the simulation of motor actions becomes object to a newly-emerging, higher-level subject. As with the first emergence of a subject-object subsystem, this will require the inhibition of immediate responses in circumstances where there is uncertainty about what adaptive response would be optimal. This suppression provides the opportunity for more effective responses to be substituted. It also enables uninterrupted internal attention to be given to the simulations by the subject-object subsystem. This in turn enables spreading activation to recruit resources that are adaptively relevant to the contents of the simulations.

Significantly, this inhibition will also provide the emerging, higher-level subject with dynamical separation from the simulations (experienced as psychological distance). This enables the new subject to begin to control the simulations, including by modifying elements within them, and assessing the consequences. This is a critically-important functional step—it enables the subject-object subsystem to change simulations/models in ways that facilitate the discovery of behaviours that may be more adaptive.

For example, the subject-object subsystem could simulate alternative motor actions and assess the salience of their impacts on the environment. This could include assessing motor actions that interact with alternative environments, including with objects and circumstances that are not in the organism's sensory fields. Eventually, this capacity could also include simulating/modelling interactions between objects in the environment, whether or not these interactions are initiated by motor actions (Kotchoubey 2018). This would enable the organism to simulate the outcome of external events in which it does not participate.

Subject-Object Emergence Theory hypothesizes that this higher-level, subject-object subsystem will be conscious of its use of its capacity to generate and manipulate simulations. In relation to real-time sensorimotor coordination, we have seen that the subject-object subsystem experiences itself as exercising voluntary choice when it consciously uses representations to coordinate its motor actions (e.g. its hand movements). For similar reasons, it will experience its manipulations of its simulations as an exercise of voluntary choice. As this modelling capacity develops, the subsystem will begin to experience itself as intentionally 'thinking through' alternative possibilities of action and assessing their consequences. The end result will be consciousness of internal simulations, and an experience of voluntary control over their modification and use.

A familiar example of the operation of this level of conscious mental processing is when we work out 'in our heads' how to accomplish a goal that requires a sequence of motor tasks e.g. the design and building of a dog kennel. We can imagine the tasks and their effects, and we can try out alternative tasks and sequences of tasks in order to check mentally whether any alternatives will contribute more effectively to the achievement of the goal. We can move backwards and forwards in our mental simulation of the sequence of tasks, exploring further alternatives as we go. These capacities can be readily confirmed by individuals who have the capacity to 'witness' the operation of their own thought processes in real time (Stewart 2007).

4.5. The adaptive significance of a capacity for concrete modelling

The acquisition of such a capacity represents a major enhancement of adaptability. It frees the organism from its previous slavery to its sensory fields. The organism can now imagine circumstances that are outside its sensory fields, including future circumstances, and take these into account as it explores adaptive possibilities. With this capacity, the dog would be able to 'see' that blocking others from their visual field does not necessarily prevent others from seeing them. Importantly, organisms at this level are able to try out possible adaptations in their heads without having to try them out in physical reality using operant or evolutionary processes. As Popper (1972) notes, this capacity "permits our hypotheses to die in our stead" (see also Dennett 1996).

Concrete-Modelling underpins tool making and tool use. It enables the organism to envisage alternative ways of using a tool that may be more effective at achieving the organism's adaptive goals. As the capacity develops, the organism will be able to simulate interactions between objects in the world. Ultimately, this enables the design and building of complicated structures, including dwellings and simple machines (Turchin 1977; Kotchoubey 2018).

Significantly, the initial operation of this capacity does not require the use of language or symbol-based reasoning. The contents of its simulations are motor actions and the impacts of the actions on the environment. The effectiveness of these simulations is not dependent on a capacity for language. Even in modern adult humans where language abilities are highly developed and have been co-opted to enhance simulation capacities, symbols do not exist for many of the components of concrete simulations (Pecher 2014). Furthermore, as can be confirmed by those with the capacity to witness their mental processes in operation, the simulation of sequences of tasks such as the building of a dog kennel do not require language capacities (see also Root-Bernstein and Root-Bernstein 1999).

As Turchin (1977) pointed out, the emergence of this capacity for internal simulation/modelling also drove the evolution of additional mechanisms that help individuals to develop and enhance the capacity. This underlines the adaptive significance of Concrete Modelling. In particular, Turchin gave as an example the propensity to engage in play which is prevalent amongst young individuals of a number of mammalian species. A key feature of play is that it fosters a capacity to imagine/simulate counter-factual social and physical circumstances, and to act as if the individual were embedded in these imagined circumstances. As well as learning which actions might be adaptive in the simulated circumstances, individuals engaging in play also develop their capacity to simulate counter-factual scenarios and to use these to explore novel adaptive possibilities.

As a second example of such a mechanism, Turchin (1977) pointed to the widespread propensity amongst humans to engage in humour. He argued that much humour involves the use of simulations/imagination to produce counter-factual and surprising juxtapositions of circumstances. Again, this fosters the development of an enhanced capacity for Concrete Modelling.

The emergence of a capacity for conscious mental modelling enabled organisms to develop models of themselves. These self-models were critically important for enabling organisms to model their interactions with their environment. This facilitated the internal modelling of alternative actions. Igamberdiev (2017) and Igamberdiev and Brenner (2020) explore how this recursive modelling process internalises representations of the external world in the self. They go on to demonstrate how the emergence of these capacities was critically important for enabling the major evolutionary transition from biological to social systems.

5. The emergence of a capacity for Abstract/Rational Modelling

5.1. The limitations of concrete modelling

I will refer to the next major milestone in the evolution/development of consciousness as the *Abstract/Rational Modelling Level*. Broadly, it encompasses some of the key cognitive capacities that underpin Piaget's *Formal Operations Stage* (Piaget 1969).

Consistent with the research strategy I have outlined, I will begin my consideration of the emergence of this new level by identifying the most significant limitations of the previous major level—the Concrete-Modelling level. As do limitations at any level, the existence of limitations at the Concrete-Modelling level provide affordances for evolutionary and developmental processes to produce higher cognitive capacities.

The key limitation at the Concrete-Modelling level is that the contents of simulations are largely restricted to concrete actions and their impacts on concrete objects and circumstances in the organism's environment. An organism at this level is largely limited to generating simulations of actions and objects that can be sensed directly. This is an obvious consequence of how Concrete Modelling originated—it emerged from simulations that represented actual motor actions, the actual sensory consequences of those actions, and the impacts of the actions on actual constituents of the organism's environment. The simulations/models were unable to deal with abstractions that could not be experienced concretely.

This is a significant limitation. Without a capacity to incorporate abstractions, simulations are largely limited to modelling particular circumstances and events. The ability to work with abstractions is essential if the organism is to discover and utilize powerful generalizations that apply across a range of particular circumstances. This ability enables an organism to recognise and make use of regularities that hold true across space and/or through time. For example, Newton's laws of motion are highly abstract and could not be developed or properly understood by cognition at the Concrete-Modelling level. As we will see in greater detail, cognitive capacities at the level of Abstract/Rational Modelling were essential to enable the full emergence and development of modern science and technology.

5.2. How further development of the subject-object subsystem could overcome limitations of concrete modelling

The research strategy I am using here raises the following issues: How could the further development of the subject-object subsystem overcome these limitations of Concrete Modelling, at least in part? More specifically, how could this be accomplished by a shift in which brain processes that previously operated in the dark, become object to consciousness? How could this enable the subject-object subsystem to develop simulations/models that are more abstract, thereby enabling the organism to escape slavery to the concrete and particular?

We have seen that at the Concrete Modelling level, the simulation of concrete actions and of concrete environmental circumstances proceeds consciously—they are object to consciousness. However, the processes that manage and manipulate these concrete models are not themselves conscious. They are components of the subject. They are developed and improved by operant learning and proceed in the dark. But these processes are prime candidates for the development of a new capacity to incorporate abstract concepts and principles into Concrete Modelling. This is because the processes that control and manage Concrete Modelling already tend to deal with these models at an abstract level—they are inherently at a meta-level to Concrete Modelling.

For example, these meta-level processes will tend to learn how to manipulate concrete models in ways that enhance the effectiveness of the models. As they learn, the meta-processes are likely to discover higher-level patterns that apply within and across models. But the capacity of the meta-processes to do this effectively is very limited until the processes become object to consciousness. The emergence of Abstract/Rational Modelling enables these processes themselves to be modelled and adapted consciously. It enables the organism to think consciously about its modelling processes and to intentionally control, adapt and enhance them. This new level is fundamentally meta-cognitive. However, it continues to include concrete modelling, albeit concrete modelling that is modified and extended.

As at previous levels, the emergence of the new level requires the inhibition of pre-existing responses. This will enable the meta-processes that manage the concrete level to be given uninterrupted attention and become object. This also creates the necessary dynamical separation between the emerging higher-level subject and the meta-processes. The separation enables the subject to manipulate the meta-processes and to intervene in their operation. This in turn enables the subject-object subsystem to simulate/model the meta-processes, and to evaluate and test alternative forms of meta-processes. With these new capacities, the subject-object subsystem is able to adapt the meta-processes so that they can be used to manage and control the concrete models more effectively. For example, an individual at this level can become consciously aware of the abstract rules and principles that constitute logical reasoning. This enables the individual to use these abstract rules consciously and intentionally to ensure that its thinking (both concrete and abstract) conforms to the dictates of rationality.

These developments were facilitated by the emergence of a capacity for symbol-based language. The ability of symbols to represent abstract concepts and categories significantly enhanced the capacity of Abstract/Rational Modelling to incorporate abstractions into its models.

5.3. The adaptive significance of a capacity for Abstract/Rational Modelling

The emergence of thinking at the Abstract/Rational level had major consequences for the adaptability and evolvability of human individuals and societies. This is the level of conscious cognition that underpinned the emergence of the European Enlightenment (Stewart 2016). As mentioned earlier, the emergence of Abstract/Rational Modelling was essential for the rise of science and technology. In particular, the proper use of the core methods of science necessitates a capacity for abstract thinking. This is obvious in the case of induction—the use of induction involves moving from specific, concrete events to hypothesizing regularities and patterns that apply across events.

But perhaps most significantly, reasoning and rationality themselves are abstract forms of thinking that cannot be developed at the Concrete-Modelling level. For example, the rules of logic are highly abstract principles that can be used to derive a wide array of specific implications from simulations and models. The emergence of Abstract/Rational thinking not only enabled the making of mental simulations and models that are more abstract, general and powerful. It also enabled the development of principles of logic and reason that could be used recursively to regulate and enhance the model-building and thinking process itself. Abstract principles of logic and rationality could be used to generate thinking and modelling that is superior at predicting how the relevant parts of reality will actually unfold. Ultimately, this facilitates superior adaptability.

6. The emergence of a capacity for Metasystemic Modelling

6.1. A new, emerging level of cognitive capacity

The next major level of conscious cognition that I will consider will be referred to as *Metasystemic Modelling.* It does not have a comparable level in Piaget's hierarchy of stages. His sequence of levels ends at the formal operations stage (broadly similar to what I refer to as Abstract/Rational Modelling). More recently, however, researchers who study adult development have identified levels beyond those described by Piaget (see, for example, Fischer, 1980; Kegan 1982; Commons et al., 1998; Mascolo 2015). Broadly, a capacity for Metasystemic Modelling underpins some of the key cognitive capacities that Commons includes in his *Metasystematic Stage* of development.

Although definitive data do not yet exist, it appears that very few humans currently operate at this emerging level. In fact, studies suggest that even in industrialised countries, as few as 30 percent of individuals attain even Piaget's formal operations level (Shayer and Wylam 1978; Pintrich 1990).

The use of Subject-Object Emergence Theory to understand the emergence of this new level of conscious cognition presents a significant opportunity for the theory. This is because the ultimate test of an evolutionary/developmental theory is whether it can predict future evolution and development. Furthermore, the development of an understanding of future evolutionary possibilities and the forces that will shape them is particularly important for organisms like humans. This is because humans have the potential to use theories of future evolution to guide and accelerate their own evolution (Stewart 2000, 2008).

6.2. The limitations of Abstract/Rational Modelling

In accordance with the research strategy outlined above, we will begin our consideration of the emergence of Metasystemic Modelling by identifying key limitations of Abstract/Rational Modelling.

We have seen that the emergence of Abstract/Rational Modelling powered the European Enlightenment, modernity, and the growth of science and technology. But despite its enormous successes, it is very limited in its ability to model aspects of reality that are complex. This in turn limits its capacity to discover effective adaptations for dealing with complex phenomena. For example, Alfred North Whitehead argued that the cognitive capacity that underpins mainstream science is unable to deal effectively with the great majority of phenomena that really matter to human beings—most of these phenomena are too complex to be understood by linear, rational, reductionist thinking (Whitehead 1925). In general, modern science has had limited success in understanding complex phenomena such as social and economic systems, ecosystems, psychology and cognition itself. When applied to these domains, current science tends to arrive at simple findings that are often trivial and fail to reflect the complexity of the phenomena.

The fundamental reason for this limitation is that the cognitive capacity that underpins much of mainstream science is incapable of adequately modelling complex systems as they evolve and interact through time (Stewart 2016). This inability of Abstract/Rational Modelling to model and to understand complex phenomena is a result of its tendency to produce reductionist, analysable models of phenomena. It produces models that can be 'thought through' using analysis and linear thinking.

The limited ability of Abstract/Rational Modelling to 'think through' complex phenomena is due to a number of factors. In particular, Abstract/Rational thinking tends to be quickly overwhelmed by increases in the number of entities that have to be represented in models (particularly given that as the number of entities increases, the number of interactions that need to be tracked and thought through tends to increase exponentially). Furthermore, such a modelling capacity tends to be overwhelmed when the phenomena being modelled are not isolated, and are being impacted continually by outside events. Abstract/Rational modelling also tends to become intractable when the causal interactions between entities in the model are not linear (particularly where circular causality is involved), and where the entities themselves transform through time—i.e. where they do not have relatively-fixed attributes.

As a consequence of these limitations, attempts by mainstream science to develop a science of complexity have tended to fall far short of what is required—as Melo (2020) argues, most attempts made to date have tended to produce mechanistic reductions of complex evolving systems. Science powered only by Abstract/Rational thinking is capable of modelling/understanding only those areas of reality that are mechanistic enough to be approximated by reductionist models that are analytically tractable. Unfortunately, most of the areas of reality that are important to human beings are not so simple.

6.3. How a capacity for Metasystemic Modelling can contribute to overcoming these limitations

Our research strategy poses the following questions: How could the further development of the subject-object subsystem overcome these limitations of Abstract/Rational Modelling, at least in part? In particular, how could this be accomplished by a shift in which brain processes that previously operated in the dark, become object to consciousness?

How could this enable humans to escape slavery to mechanistic, analytical, logical thinking that fails to reflect the complexity of much of reality?

Like the transition to Abstract/Rational Modelling, the shift to Metasystemic Modelling requires that the processes that adapt and operate Abstract/Rational Modelling become object to consciousness. Previously, these processes operated in the dark, and were shaped by associative and operant learning processes. But with the emergence of a new, higher-level subject to which these processes are object, the organism can develop the capacity to consciously control and adapt its Abstract/Rational Modelling. Operating at a meta-level to Abstract/Rational Modelling, the organism can think consciously about the methods it uses to construct and operate its Abstract/Rational Modelling. This enables the new subject to begin to 'see' the limitations of its current modelling, and begin to develop modelling capacities that are better able to represent and model complex aspects of reality. The emerging subject-object subsystem will begin to experience itself as having voluntary control over its Abstract/Rational Modelling capacity.

Basseches (1984) and Laske (2009) identify a range of thought processes that an individual would need to develop in order to overcome the limitations of Abstract/Rational Modelling and to enable what I term Metasystemic Modelling. They refer to these thought processes as 'movements in thought' and 'thought forms'. Laske classified these into four quadrants. Each quadrant represents a class of thought forms that is not adequately represented in mechanistic, reductionist Abstract/Rational Modelling. Metasystemic Modelling needs to incorporate these four classes of thought forms if it is to be capable of representing and understanding complex reality.

The four quadrants comprise: (1) the *process quadrant*, which contains thought forms that recognise the fact that reality is ceaselessly and continuously changing—all 'objects' are in fact reified processes. All objects have a history of past transformation, and will continue to transform indefinitely into the future; (2) The *context quadrant* which contains thought forms that recognise that no phenomenon is isolated from its environment—all phenomena are embedded in a multi-level hierarchy of phenomena that interact with them; (3) the *relationship quadrant*, which recognises that processes within a system tend to coevolve in relationship with other processes in the system; and (4) The *transforming systems quadrant*, which reflects the fact that when the other three quadrants are properly taken into account, all phenomena can be seen to be participants in systems that interact and transform through time.

6.4. *The need to incorporate additional psychological resources in order to enable complex mental modelling*

The conscious incorporation of these *movements in thought* into the emerging subject-object subsystem goes some way towards enabling Metasystemic Modelling. But by itself it is not enough. This is because thinking, including the symbol-based thinking that increasingly dominates Abstract/Rational Modelling as it develops, has significant limitations. In particular, it is not very effective at representing and manipulating patterns, including complex patterns of relationships in social and other systems.

In order to overcome this limitation, the full development of Metasystemic Modelling requires that other psychological resources need to become object to the emerging subject, and be incorporated into the new modelling capacity. These psychological resources include those associated with pattern-recognition, intuition, images and feelings. The incorporation of these resources can produce modelling that is able to deal with phenomena that are unable to be analysed and thought through. It facilitates mental modelling that is not dominated by the use of propositions and logic. To deal adequately with complex phenomena, our mental models need to know far more than we can tell.

For example, our emotional systems are often very effective at recognising and appraising the patterns involved in complex social situations. Our emotions can instantly assess the import of social circumstances that are highly complex. In contrast, the use of logical, thought-based skills to recognise and evaluate what is going on in social situations is notoriously ineffective (McGilchrist 2009). In part, this is why some people with autistic tendencies have difficulty negotiating complex social circumstances, even though they might be highly competent at using language-based analysis to model mechanistic phenomena (Baron-Cohen 1995).

However, until the emergence of the Metasystemic level of development, these resources and functions tend to operate outside of consciousness. They are established and adapted by evolutionary processes and by learning. Largely, they operate in the dark.

If these additional resources as well as the processes that regulate Abstract/Rational Modelling are to become object to consciousness, the emerging subject-object subsystem needs to be able to inhibit pre-existing responses, particularly where there is uncertainty. As was the case in earlier transitions, this enables uninterrupted attention to be given to the processes that are newly becoming object. As well as enabling the processes to become object to the new subject, it also enables alternative responses to be substituted. Furthermore, it also produces the dynamical separation between subject and object that enables the subject to control these processes. This is experienced as providing psychological distance from the processes, and voluntary control over them. The individual is no longer embedded in Abstract/Rational Modelling.

6.5. *The adaptive significance of a capacity for Metasystemic Modelling*

The full attainment of a capacity for Metasystemic Modelling enables individuals to intentionally build mental models of multi-layered complex systems as they interact and evolve through time. Individuals can use their voluntary control over the models to mentally compare and contrast systems, to mentally modify processes within the models and assess the consequences, and to mentally model alternative interventions in the models in order to identify the interventions that would contribute most to the achievement of their adaptive goals.

The use of digital simulations is not a substitute for Metasystemic Modelling, although simulations can assist mental modelling. Like complex aspects of reality, complex digital simulations can only be understood and appropriately manipulated by individuals with a capacity for Metasystemic Modelling. In significant part, this is why access to computer simulations has not opened the door to a genuine science of complexity.

The widespread emergence of Metasystemic Modelling will have major consequences for the adaptability and evolvability of humans, as did the emergence of Concrete Modelling and Abstract/Rational Modelling before it. Abstract/Rational Modelling powered the rise of modern science and technology. But as we have seen, science underpinned by Abstract/Rational Modelling has enabled only a small proportion of reality to be modelled and understood. Metasystemic Modelling will enable science to model much of the remainder of reality that is too complex to be modelled effectively by reductionist, analytical thinking.

Currently, it is rarely recognised that modern abstract/rational science is as incapable of understanding complex phenomena as dogs are incapable of understanding and taking into account events outside their sensory field. Individuals at a particular level of cognitive capacity are incapable of seeing what is missed at their level. This is because they are incapable of mentally modelling what is left out at that level—to be able to do so would require cognitive capacities that are at least at the next highest level.

Nevertheless, some of the best scientific minds of the 20th century have drawn attention to the inability of modernist science to deal with complex phenomena, and have attempted to initiate research programs to develop new kinds of science that would overcome this. Examples include Holism (Smuts 1926; Blitz 1992); Cybernetics and the Macy

Conferences (McCulloch 1974); General Systems Theory (Von Bertalanffy 1968); Self-Organisation (Salthe 1985; Heylighen 2001); the Santa Fe Institute and Complexity Science (Mitchell 2009); Complex Adaptive Systems (Miller and Page 2007) and universal Evo-Devo Theory (Vidal 2010). However, none of these initiatives have yet succeeded in igniting the scientific revolution that will follow the successful and widespread extension of science into complex domains.

In large part, this is because the criteria used by Abstract/Rational science to assess the validity of scientific theories and research programs tend to rule out approaches that are powered by Metasystemic Modelling. Abstract/Rational science demands analytical, logical and empirical 'rigor'. This is achievable for phenomena that are simple and mechanistic. But it is often impossible for complex phenomena. As a result, attempts to develop a science of complex, evolving, multilayered phenomena can generally be shown by mainstream science to fail the central requirements of its kind of science.

To date, attempts to break out of these inappropriate strictures have not proven successful. In significant part this is due to the fact that very few scientists have yet developed a strong capacity for Metasystemic Modelling. Abstract/Rational cognition prevails overwhelmingly in current mainstream science.

Science underpinned by Metasystemic Modelling has been able to escape rejection by the mainstream only in limited circumstances. For example, it has avoided rejection where it has used its complex models to derive narrow, analytically-tractable results that can survive testing against Abstract/Rational criteria. But it has not been able to gain acceptance from the mainstream for the complex mental models that were actually used to generate these narrow results. This is why most of what currently passes for complexity science is in fact a mechanistic reduction of complex processes and systems. It is only these kinds of mechanistic reductions that can satisfy the requirements of mainstream science. Mainstream science is currently a slave to analytical, logical, reductionist thinking and its associated methodologies.

7. The emergence of Voluntary Emotional Control

7.1. The limitations of pre-existing emotional capacities

The final subject-object transition that I will consider occurs when emotions and associated behaviours become object. I will refer to this transition as the shift to Voluntary Emotional Control. This transition enables our emotional systems to be controlled consciously by the subject-object subsystem. Until this transition is accomplished, the emotional system operates largely outside conscious awareness, in the dark (but this is not to say that individuals at lower levels are unaware of all aspects of their emotions—for example, they can be conscious of some of the sensory impacts of emotional processes). At lower levels, our emotions have us, we do not have emotions. We are embedded in our emotions, and our emotional responses tend to be outside our voluntary control. This locks us into learned and innate emotional responses that might not be optimal in particular circumstances (Kegan 1982; Stewart 2000, 2007). The development of a capacity for Voluntary Emotional Control enables individuals to escape this slavery to their pre-existing emotional responses and motivations.

7.2. How the limitation of previous levels can be overcome by Voluntary Emotional Control

As for previous transitions, the shift to Voluntary Emotional Control requires that relevant processes in the emotional system become object to consciousness. As we have seen, this requires the subject-object subsystem to develop the capacity to inhibit learned and innate responses that are evoked by emotions. This will enable emotions themselves to be given uninterrupted attention. It also enables alternative responses to be developed using conscious modelling processes, and for these alternative responses to be enacted instead. The newly-emerging subject experiences this as having some psychological distance from its emotions and being able to exercise voluntary control over its emotional responses.

The emergence of Voluntary Emotional Control can occur independently of cognitive development—in principle, it can occur while the individual is at any of the three levels of cognitive modelling that I have outlined above. However, as the cognitive capacity of individuals develops through these levels, they are increasingly likely to see the desirability of consciously controlling their emotions and associated behaviours. This is because their improving modelling capacity is increasingly able to identify circumstances in which their learned and inherited emotional responses are sub-optimal.

Humans have developed an array of techniques for intentionally promoting the development of Voluntary Emotional Control. These techniques involve the use of practices that tend to shift emotional processes from subject to object. Central to these practices are various forms of meditation (e.g. see Tart 1987 and also the model of meditation developed by Lefebvre 2017). From the perspective being developed here, appropriate meditation practices have the potential to make object processes that previously operated in the dark (although meditation is often used for other purposes entirely). When individuals first begin to use these practices, they experience themselves as being almost continuously embedded in thinking and emotions. When this is the case, these processes are object to consciousness only to a limited extent.

However, the core element of many meditation practices can train a capacity to dis-embed from thought and emotion. This core element involves the repeated practice of dis-engaging attention from embeddedness, and moving surrendered, 'bare' attention to sensations. Resting attention on sensations makes the sensations object to consciousness, and inhibits alternative responses. In particular, thoughts and feelings that might re-embed awareness tend to be inhibited. Repetition of this core practice builds a capacity to dis-engage the individual from inherited and habitual responses. It also trains the ability to give dis-embedded, bare attention to thoughts and emotions as they arise i.e. they also become object to consciousness. With continued practice, the individual can develop the capacity to remain dis-embedded from thought and emotion for extended periods. Eventually, they can learn to do this in real time in the midst of ordinary life, as thoughts and emotions arise (for a detailed information-processing theory of meditation that examines these issues in depth, see Stewart 2007).

It is not surprising that the founders of the world's great spiritual and religious traditions strongly promoted the use of meditation practices. This is because of the potential of meditation to build a capacity for Voluntary Emotional Control. Such a capacity is essential if the followers of the traditions are ever to live in accordance with the behavioural injunctions established by the founders e.g. to 'resist temptation' and 'turn the other cheek' (Christianity), to free oneself from all desires (Buddhism), to experience equanimity in the face of pleasure or pain (Hinduism), and to transcend the self-centred desires and grasping that underpin ego (common to many traditions) (Stewart, 2017). However, the forms of meditation that are most effective at producing capacities for Voluntary Emotional Control tend to have been lost by these traditions, and their modern adherents are rarely capable of actually following these injunctions.

By scaffolding the capacity to make thought processes object to consciousness, appropriate meditation practices also have the potential to assist the development of higher levels of cognition (Stewart 2007, 2017). However, meditation practices have been used for this purpose to a much lesser extent than for emotional development. In large part, this is because limitations in cognition are far less easy for individuals to see compared with limitations in emotional responses. As we have discussed, this is because individuals at a particular cognitive level tend to be blind to the limitations of cognition at that level. This manifests as a lack of motivation to intentionally develop higher cognition. However, appropriately-designed meditation practices have the potential to scaffold the expansion of conscious control in all domains, including in cognitive development.

7.3. *The adaptive significance of a capacity for Voluntary Emotional Control*

The evolutionary and developmental advantages of the transition to Voluntary Emotional Control are substantial—it has the potential to significantly enhance adaptability. The transition enables mental modelling to be used to review whether learned and innate emotional responses are optimal in specific circumstances. In particular, it enables the individual to use mental modelling to assess whether superior adaptive responses are available. This is a particularly significant capacity for humans because our innate responses have been shaped by evolution in past environments, and may not be optimal in current circumstances.

Furthermore, as individuals grow and develop, and as their modelling and other cognitive capacities are enhanced, they are often capable of identifying adaptations that are superior to those they learned at younger ages, including during their childhood. As their cognition develops, they increasingly realize that adaptations learned using inferior learning processes often need to be updated. This realization is particularly marked when the individual reaches the Metasystemic Modelling level. This is because it is the first level of conscious modelling that can discover effective adaptations in the complex circumstances that are often dealt with by our emotional systems.

In general, Voluntary Emotional Control enables humans to free themselves from the dictates of their evolutionary, social and cultural past, and to instead adapt in ways that serve the demands of current and future evolution (Stewart 2008). Stewart (2000) refers to individuals who have developed this capacity as 'self-evolving organisms', and suggests that their emergence constitutes a major evolutionary transition in evolvability.

As the conscious subsystem evolves and develops, it extends conscious control to an increasing proportion of brain processes. When the minimally-complex sensorimotor subsystem first emerges, it is just a tiny spark of light in an ocean of darkness. But from there it progressively expands. As conscious modelling develops, it is increasingly able to adapt the organism more effectively than pre-existing inherited and learned processes. As discussed, the improving modelling capacity can be used to revise the adaptations that were previously embodied in emotional and intuitional systems. More generally, as the conscious subsystem evolves and develops, the procedural knowledge that was embodied in learned and innate processes is increasingly translated into mental models that utilize declarative knowledge. At the evolutionary level, this is equivalent to the 'procedural to declarative re-description' which is identified by Karmiloff Smith (1992) as occurring as humans develop (see also Stewart 2007).

However, this does not mean that these pre-existing emotional and other capacities are abandoned. As we have discussed, emotional and intuitional processes contain pattern-recognition and other resources that cannot be replaced easily by thought-based modelling. At the Metasystemic Modelling level, the development of conscious control over emotional and intuitional systems enables these irreplaceable resources to be integrated into the cognitive modelling process itself.

Nor does this reconfiguring of the emotional system mean that emotions are suppressed or repressed. As with all processes that become object to consciousness, they light up and are experienced more vividly. The key change is that it becomes a matter of conscious choice whether the responses that were previously evoked automatically are now acted upon. The making of these conscious choices is informed by conscious mental modelling.

8. Testing the Subject-Object Emergence Theory

8.1. *The theory makes numerous testable predictions*

Subject-Object Emergence Theory makes many bold and novel predictions that are testable and potentially falsifiable. In particular, it makes numerous specific predictions about how subsystems that give rise to conscious experience are organised functionally. Furthermore, the hypothesis identifies specific circumstances in which it predicts that brain processes that previously operated in the dark will become object to a newly-emerging subject, and will be experienced consciously. These circumstance can arise during evolution, individual development, and through the use of meditative practices The hypothesis makes specific predictions about the nature of the processes that will need to emerge (including how they will need to interrelate and function), in order to produce these shifts from processing in the dark to conscious experience.

In particular, the theory posits that consciousness in biological organisms arises through the emergence of subject-object subsystems. Accordingly, it predicts that all instances of conscious experience will be found only where an appropriate subject-object subsystem exists and functions. This will be the case whether a particular instance of consciousness is already established in an organism, or whether it arises during development. The theory predicts that these subject-object subsystems will be found to be constituted by specific processes (these are identified in Section 3 above). It also predicts that if particular components of a subject-object subsystem are prevented from functioning, the corresponding conscious experience will not be produced.

Furthermore, many of the key predictions of the theory can be tested by individuals who have developed the capacity to witness their own psychological functioning and development in real time (for details about the development of this capacity, see Stewart 2007).

It is obviously beyond the scope of this paper to develop a detailed research program that will test the theory comprehensively. However, we will now briefly consider some key areas in which specific tests would be particularly powerful for assessing the theory, testing it against competing theories of consciousness, and refining the theory.

8.2. *Competing hypotheses –The Hard Problem*

The central argument advanced by Chalmers (1996) in his discussion of The Hard Problem is that a physical reduction of consciousness is impossible. More specifically, he argues that it is impossible to demonstrate that any particular set of physical brain processes will necessarily give rise to conscious experience—physical brain processes alone cannot be shown analytically to produce non-physical phenomena such as consciousness. He concludes that it cannot be proven deductively that any set of physical brain processes will feel like something to have, rather than just function in the dark.

If the only strategy available to science to understand consciousness necessitated such a physical reduction, Chalmers' convincing (but limited) analysis would mean that consciousness is indeed a very hard problem for science. But it is not. Chalmers rules out only one narrow, analytic approach. He does not consider other approaches available to science that have been far more successful in dealing with similar challenges in other domains.

In general, approaches that rely on the development of novel, falsifiable hypotheses that are subjected to rigorous testing allow science to pursue a much broader range of research strategies (Popper 1959; Solms, 2021; Thornton 2021). Science is not restricted to considering only hypotheses that enable the existence of conscious experience to be deduced from physical brain processes alone. In particular, it is allowable within science to test hypotheses that postulate that particular physical brain processes produce conscious experience, even though it cannot be demonstrated analytically that the particular physical processes alone can produce this experience. As we will discuss further below, it is commonplace in science to consider hypotheses that conjecture that particular physical processes give rise to novel phenomena that are qualitatively very different to the attributes of those physical processes. In relation to consciousness, it is equally appropriate for science to consider hypotheses that posit that particular physical processes produce consciousness even though consciousness has attributes that are very different to the attributes of the physical processes and cannot be deduced from them. Whether such hypotheses should be rejected or considered further is a matter to be determined by the outcome of attempts to test and falsify their predictions empirically.

The potential of such approaches has been demonstrated in many other areas of science where properties cannot be deduced from underlying processes and where narrow, reductive approaches fail. For example, it is often impossible to deduce the nature of the properties and attributes of complex systems from models of the internal constituents of the systems alone. This is most obvious where properties and attributes of the systems are emergent phenomena. But the failure in these circumstances of analytic approaches that rely only on physical reduction has not prevented science from developing and testing falsifiable hypotheses about the processes that produce the properties. It is entirely consistent with the methods of science to hypothesise that particular physical processes give rise to novel and surprising properties that are not deducible from their constituent processes.

For example, science has made considerable progress in understanding the novel properties of molecules that are not possessed by their constituent atoms. A further example concerns the development of an understanding of the properties that distinguish life. A century ago, the task of explaining life was seen by many as a hard problem, for similar reasons to those given by analytic philosophers in relation to consciousness. Living systems have attributes that cannot be deduced from an understanding of their physical constituents alone. But the broader processes of science have essentially dissolved this problem, as a number of consciousness researchers have noted (e.g. see Klein and Barron 2020). Science has achieved this through broad research programs that have progressively accumulated and refined hypotheses that survive attempts to falsify them.

Where empirical science is founded on such approaches, it progresses through the accumulation of hypotheses that have not yet been falsified. It is not restricted to the accumulation of hypotheses that are ‘proven’ and deduced analytically.

For these reasons, Chalmers’ arguments do not provide any a priori reason to believe that the broad research program advanced by this paper cannot progressively develop, test and refine hypotheses that explain consciousness.

8.3. *Competing hypotheses – Global Workspace Theory (GWT)*

There are a number of testable differences between the Subject-Object Emergence Theory and GWT, which is arguably the most widely accepted functionalist approach to consciousness. The functional architecture postulated by the two theories differs in significant ways. I will briefly discuss here one such difference that leads to very different predictions about the nature of the specific functioning that produces conscious experience.

GWT hypothesizes that a global broadcast process is an essential component of the brain processes that give rise to consciousness (Baars et al., 2013). In contrast, the Subject-Object Emergence Theory hypothesizes that consciousness arises due to the emergence of a subject-object subsystem, whether or not the subsystem includes processes that entail global broadcasting. More specifically, it hypothesizes that the emergence of a minimally-complex subject-object subsystem will give rise to conscious experience. It predicts that this will be the case even though such a subsystem may emerge *before* the emergence of global broadcasting processes. More specifically, it predicts that conscious experience can emerge before the existence of the neuronal infrastructure that is necessary to instantiate a global workspace and associated broadcasting (e.g. before the emergence of the long-distance neural circuitry outlined by Dehaene and Naccache 2001). The subsequent emergence of global broadcasting could be expected to significantly improve the adaptability of the organism. However, it would evolve/develop after the emergence of a subject-object subsystem and associated conscious experience.

8.4. *Instantiation in AI*

In principle, it should be possible to instantiate the emergence of a minimally-complex subject-object subsystem in suitable AI. This would provide critically important opportunities for testing the Subject-Object Theory. However, it would require instantiation rather than a digital simulation. A digital simulation of subject-object emergence would not be expected to produce conscious experience any more than a digital simulation of a cyclone would be expected to produce actual rain or wind. Or that a digital simulation of an organism would be expected to be actually alive.

A far more complex challenge for AI research would be to instantiate AI that moves through developmental levels similar to those that have been identified in this paper. Instead of its development ceasing with the emergence of a minimally complex subsystem, this AI would proceed to develop through the kind of levels that characterize human cognitive and social/emotional development. This would be likely to necessitate structures and systems that scaffold the AI’s development through the sequence of levels. The scaffolding would need to include appropriate analogues of socialisation, education, cultural interactions (including analogues of play and humour), and meditation practices.

9. Conclusion

This paper demonstrates that an evolutionary/developmental research strategy is capable of producing a rich set of novel and powerful predictions about the functioning of consciousness. Central to this research strategy is a focus on circumstances in which processes that previously operated in the dark become object to a newly-emerging subject. These transitions can occur during the development of individuals as well as during evolution. The key challenges facing the research strategy are to identify how brain processes that emerge during these transitions are organised functionally, how their functional organisation differs from the organisation of pre-existing processes that operated in the dark, and how these differences produce new classes of conscious experience.

The paper began its implementation of the evolutionary/developmental research strategy by attempting to reconstruct the simplest form of subject-object subsystem that might first emerge and give rise to conscious experience. This reconstruction was informed by the necessity for the minimally-complex subsystem to provide adaptive benefits if its emergence were to be favoured by evolutionary/learning processes.

The application of this strategy generated a novel hypothesis that makes numerous testable predictions. The hypothesis suggests that the emergence of consciousness was driven by the ability of a minimally-complex, subject-object subsystem to enable sensorimotor coordination in real time. More specifically, such a subsystem could enable the direction of attention to be coordinated with salient features in the environment, in real time. A more familiar but more complex example of sensorimotor coordination is real-time hand-eye coordination.

Importantly, the minimally-complex subject that emerges with this new subsystem is simple enough to be understood functionally. This functional understanding does not have to resort to the use of ‘black boxes’ that do the ‘hard work’ of explaining consciousness. As such, it does not commit the homunculus fallacy.

The Subject-Object Emergence Theory hypothesizes that the subsystem that implements this sensorimotor coordination will experience its actions as voluntary, and will be conscious of the representations it uses to coordinate its actions.

The paper then continues its application of the evolutionary/

developmental research strategy by setting out to reconstruct subsequent major transitions in the subject-object subsystem. In each transition, additional processes that previously operated in the dark became object to consciousness for the first time. A key step in each reconstruction is to identify the limitations in adaptability that manifest at the previous level. These limitations are affordances that drive the emergence of a new level that is capable of overcoming at least some of the limitations.

The application of this strategy leads to the reconstruction of a sequence of transitions that involve the development of a capacity for conscious mental modelling. The initial significance of such an ability is that it enables organisms to escape slavery to their sensory fields. It enables organisms to take into account circumstances that are not represented in their sensory fields, but are represented in their mental models. For example, it enables organisms to 'go offline' from their sensory fields and to use models/simulations to predict how their environment will be impacted by possible actions. They can then use these predictions to identify actions that will be adaptive.

Building on a hypothesis advanced by Pezzulo (2011), the paper suggests that such an ability for mental modelling emerged from a pre-existing capacity that operated initially as part of the organism's Predictive Processing capabilities. This capacity enabled organisms to simulate motor actions in order to anticipate sensory and other consequences of their actions. As the capacity developed, elements of the simulations could be changed, and the consequences of these changes could be thought through consciously, enabling alternative adaptations to be evaluated.

The final cognitive transition considered by the paper is the emergence of Metasystemic Modelling. Currently, this capacity has barely begun to arise amongst humans. The paper argues that the spread of Metasystemic Modelling will enable the mental modelling of complex, evolving phenomena. This in turn will make possible a new kind of science. Its emergence is predicted to be as significant for humanity as was the spread of a capacity for Abstract/Rational Modelling which drove the European Enlightenment.

As well as providing numerous testable predictions, the research strategy outlined by the paper also has the potential to be used to engineer and train AI that is conscious. Furthermore, it has the potential to generate approaches that could accelerate the development of a capacity for Metasystemic Modelling in both humans and AI.

Declaration of competing interests

None.

Data availability

No data was used for the research described in the article.

Acknowledgements

The author acknowledges the value of useful discussions and comments from Stefan Pernar, William Davis and Delena Vo.

References

Baars, B.J., 1988. A Cognitive Theory of Consciousness. Cambridge University Press, Cambridge.

Baars, B.J., Franklin, S., Ramsoy, T.Z., 2013. Global workspace dynamics: cortical "binding and propagation" enables conscious contents. Front. Psychol. 4, 200.

Basseches, M., 1984. Dialectical Thinking and Adult Development. Ablex, Norwood, NJ.

Baron-Cohen, S., 1995. Mindblindness: an Essay on Autism and Theory of Mind. MIT Press, Cambridge, MA.

Bertenthal, B.I., 2020. Motor experience and action understanding: A developmental systems perspective. In: Bidell, T., Mascolo, M.F. (Eds.), Handbook of integrative psychological development: Essays in honor of Kurt W. Fischer. Taylor and Francis, Abingdon, UK.

Blitz, D., 1992. Emergent Evolution: Qualitative Novelty and the Levels of Reality. Kluwer Academic Publishers, Dordrecht.

Chalmers, D.J., 1996. The Conscious Mind. Oxford University Press, New York.

Combs, A., 2009. Consciousness Explained Better. Paragon House, St. Paul, MN.

Commons, M.L., Trudeau, E.J., Stein, S.A., Richards, F.A., Krause, S.R., 1998. Hierarchical complexity of tasks shows the existence of developmental stages. Dev. Rev. 18, 237–278.

Dehaene, S., Kerszberg, M., Changeux, J.A., 1998. A neuronal model of a global workspace in effortful cognitive tasks. Proc. Natl. Acad. Sci. Unit. States Am. 95, 14529–14534.

Dehaene, S., Naccache, L., 2001. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition 79, 1–37.

Dennett, D.C., 1991. Consciousness Explained. Little Brown, New York.

Dennett, D.C., 1996. Kinds of Minds. Basic Books, New York.

Fischer, K.W., 1980. A theory of cognitive development: the control and construction of hierarchies of skills. Psychol. Rev. 87, 477–531.

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11, 127–138.

Graziano, M.S.A., Webb, T.W., 2016. From sponge to human: the evolution of consciousness. In: Kaas, J. (Ed.), Evolution of Nervous Systems, second ed., vol. 3. Elsevier, New York, pp. 547–554.

Gunji, Y.P., Shinohara, S., Haruna, T., Basios, V., 2017. Inverse Bayesian inference as a key of consciousness featuring a macroscopic quantum logical structure. Biosystems 152, 44–65.

Hawkins, J., 2004. On Intelligence. Times Books, New York.

Heylighen, F., 2001. The science of self-organization and adaptivity. Encycl. Life Support Syst. 5, 253–280.

Heylighen, F., Bollen, J., 1996. The world-wide web as a super-brain: from metaphor to model. In: Trappl, R. (Ed.), Cybernetics and Systems '96. Austrian Society for Cybernetics, pp. 917–922.

Igamberdiev, A.U., 2017. Evolutionary transition from biological to social systems via generation of reflexive models of externality. Prog. Biophys. Mol. Biol. 131, 336–347.

Igamberdiev, A.U., Brenner, J.E., 2020. The evolutionary dynamics of social systems via reflexive transformation of external reality. Biosystems 197, 104219.

Karmiloff-Smith, A., 1992. Beyond Modularity: A Developmental Perspective on Cognitive Science. MIT Press, Cambridge MA.

Kegan, R., 1982. The Evolving Self: Problems and Process in Human Development. Harvard University Press, Cambridge, MA.

Kegan, R., 1994. In over Our Heads: the Mental Demands of Modern Life. Harvard University Press, Cambridge, MA.

Klein, C., Barron, A.B., 2020. How experimental neuroscientists can fix the hard problem of consciousness. Neurosci. Conscious. 1, niaa009.

Kotchoubey, B., 2018. Human consciousness: where is it from and what is it for. Front. Psychol. 9, 567.

Laske, O.E., 2009. Measuring hidden dimensions. In: Foundations of Requisite Organization, vol. 2. Interdevelopmental Institute Press, Medford, MA.

Lefebvre, V.A., 2017. Theoretical modeling of the subject: Western and Eastern types of human reflexion. Prog. Biophys. Mol. Biol. 131, 325–335.

Mascolo, M., 2015. Neo-Piagetian theories of cognitive development. In: Wright, J. (Ed.), International Encyclopedia of the Social & Behavioral Sciences, second ed. Elsevier, New York, pp. 501–510.

Mashour, G.A., Roelfsema, P., Changeux, J.P., Dehaene, S., 2020. Conscious processing and the global neuronal workspace hypothesis. Neuron 105, 776–798.

McCulloch, W.S., 1974. Recollections of the many sources of cybernetics. Am. Soc. Cybern. Forum 6 (2), 5–16.

McGilchrist, I., 2009. The Master and His Emissary: the Divided Brain and the Making of the Western World. Yale University Press, New Haven and London.

Melo, A.T., 2020. Performing Complexity: Building Foundations for the Practice of Complex Thinking. Springer International Publishing, New York.

Miller, J.H., Page, S.E., 2007. Complex Adaptive Systems: an Introduction to Computational Models of Social Life. Princeton University Press, Princeton, NJ.

Mitchell, M., 2009. Complexity: A Guided Tour. Oxford University Press, Oxford, UK.

Noë, A., O'Regan, J.K., 2000. Perception, attention and the grand illusion. Psyche 6, 15.

Pecher, D., 2014. The role of motor affordances in visual working memory. In: The Baltic International Yearbook of Cognition, Logic and Communication, vol. 9. New Prairie Press, Kansas.

Pezzulo, G., 2011. Grounding procedural and declarative knowledge in sensorimotor anticipation. Mind Lang. 26, 78–114.

Piaget, J., 1969. The Psychology of the Child. Basic Books, New York.

Pierson, H.A., Gashler, M.S., 2017. Deep learning in robotics: a review of recent research. Adv. Robot. 31 (16), 821–835.

Pintrich, P.R., 1990. Implications of psychological research on student learning and college teaching for teacher education. In: Houston, W.R. (Ed.), Handbook of Research on Teacher Education. McMillan, New York, pp. 826–857.

Popper, K.R., 1959. The Logic of Scientific Discovery. Hutchinson, London.

Popper, K.R., 1972. Objective Knowledge - an Evolutionary Approach. Clarendon, Oxford.

Root-Bernstein, R., Root-Bernstein, M., 1999. Sparks of Genius: the Thirteen Thinking Tools of Creative People. Houghton, Mifflin and Company, New York.

Salthe, S.N., 1985. Evolving Hierarchical Systems. Columbia University Press, New York.

Shayer, M., Wylam, H., 1978. The distribution of Piagetian stages of thinking in British middle and secondary school children II: 14-16 year olds and sex differentials. Br. J. Educ. Psychol. 48, 62–70.

Shen, G., Dwivedi, K., Majima, K., Horikawa, T., Kamitani, Y., 2019. End-to-End deep image reconstruction from human brain activity. Front. Comput. Neurosci. 13, 21.

Skinner, B.F., 1981. Selection by consequences. Science 213 (4507), 501–504.

Smuts, J.C., 1926. Holism and Evolution. The Macmillan Company, New York.

Solms, M., 2021. The Hidden Spring: A Journey to the Source of Consciousness. Profile Books, London.

Stewart, J.E., 2000. Evolution's Arrow: the Direction of Evolution and the Future of Humanity. The Chapman Press, Canberra.

Stewart, J.E., 2007. The future evolution of consciousness. J. Conscious. Stud. 14, 58–92.

Stewart, J.E., 2008. The evolutionary manifesto. http://www.evolutionarymanifesto.com/man.pdf.

Stewart, J.E., 2016. Review of dialectical thinking for integral leaders: a primer, by otto Laske. Integr. Leader Rev. http://integralleadershipreview.com/14809-14809/.

Stewart, J.E., 2017. Enlightenment and the evolution of the material world. Spanda J. VII, 107–114, 1.

Tart, C., 1987. Waking up. Shambhala, Boston.

Thornton, S., 2021. Karl Popper. In: Zalta, N. (Ed.), The Stanford Encyclopedia of Philosophy (Fall 2021 Edition), URL: https://plato.stanford.edu/archives/fall2021/entries/popper/.

Turchin, V.F., 1977. The Phenomenon of Science: A Cybernetic Approach to Human Evolution. Columbia University Press, New York.

Unger, P., 2014. Empty Ideas: A Critique of Analytic Philosophy. Oxford University Press, Oxford.

Vidal, C., 2010. Introduction to the special issue on the evolution and development of the Universe. Found. Sci. 15 (2), 95–99.

Von Bertalanffy, L., 1968. General System Theory: Foundations, Development. George Braziller, New York.

Vygotsky, L.S., 1978. Mind in Society: the Development of Higher Psychological Processes. Harvard University Press, London.

White, B., 2021. The hard problem isn't getting any easier: thoughts on Chalmers' "meta-problem". Philosophia 49, 495–506.

Whitehead, A.N., 1925. Science and the Modern World [1967. Free Press, New York.